Stationary models using latent structures

Mena-Chavez, Ramses H.

Award date:
2003

Awarding institution:
University of Bath

Stationary Models Using Latent Structures

submitted by

Ramsés H. Mena-Chávez

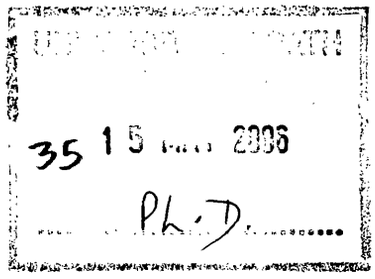
for the degree of Ph.D.

of the

University of Bath

2003

ProQuest LLC
789 East Eisenhower Parkway
P.O. Box 1346
Ann Arbor, MI 48106-1346


35 15 MAY 2006
Ph.D.

To Ruth

Abstract

Stationary models arise in a wide variety of problems where statistical analysis is applied. They naturally occur when considering phenomena which evolve over time and retain certain distributional features. In particular, it is of interest to construct stationary models with invariant distributions belonging to specific parametric families.

This thesis considers novel approaches to construct and represent stationary Markov models with given stationary distributions. We explore the use of latent structures to perform such tasks. The construction through latent structures simplifies some issues like estimation and simulation of the models. An idea introduced by Pitt et al. (2001) is studied and generalized. Specifically, some stationary models, in both discrete and continuous time are presented. Examples of these models are ARCH-type models and some diffusion processes.

Another approach to construct stationary models with specific invariant distributions is based on self-decomposable random variables. This approach has been recently used by Barndorff-Nielsen and Shephard (2001) to construct volatility models. In this thesis, an alternative approach, also based on a latent representation, is presented.

Acknowledgements

This Thesis would not have been possible without the help of many people and entities.

First, my supervisor Professor Stephen G. Walker whose friendship, support and advice were essential during the course of my Ph.D.

I am also indebted to Dr. Mogens Bladt, and all my teachers from the UNAM who guided and encouraged my career.

All the people from the Department of Mathematical Sciences at the University of Bath, for their everyday help, specially the secretaries, my officemates and computing support.

Lastly, I wish to thank the *National Council for Science and Technology* (CONACYT), México, for its financial support to pursue this work.
Scholarship 149103/154941.

Contents

CHAPTER 1

INTRODUCTION

The uncertainty present in many phenomena can be modelled with stochastic processes. Typically, the nature of the phenomenon determines the modelling alternatives to be used. That is, the specification of the different features which define the associated stochastic process. Those models presenting similar distributional features at different time points will be of particular interest in this thesis. Specifically we will concentrate on models where certain marginal properties remain invariant as time passes. These phenomena can be found in many applications, for example, short interest rates are known to have this behavior. An important class of stochastic processes encompassing the above mentioned models is the class of stationary processes.

Classic theory of stochastic processes provides with ways of finding conditions for stationarity on more general models. However, under some modelling contexts, it may be of interest to restrict only to stationary models. If, for example, we know from the outset that the phenomenon under study cannot have an explosive behavior, then imposing a model that allows for explosions may require further restrictions. For some models, such as Gaussian processes, well-known conditions for stationarity are available, however in more complicated situations this is not always the case.

An alternative approach is to directly construct stationary processes fitting the given marginal information, rather than finding conditions on more general settings. This approach is relatively new and has been introduced in different areas, for different purposes and with different names, references are given in Chapter 3. For instance, when the model is time-homogeneous then one can speak of a model with given marginals or given invariant distributions, which in turn implies that the model is stationary. Given

a particular problem or phenomenon, the specification of the stationary distribution may not be a trivial issue. In most situations, this can be given up to a parametric family. Under some modelling contexts, such specification is based on empirical results or model assumptions. In this work, we assume that the specification of the parametric family to which the theorized invariant distribution belongs has been fixed.

Once a suitable parametric family, containing the invariant distribution, has been chosen the problem is to construct a suitable model that represents the data. Notice that setting such invariant distribution does not necessarily imposes unique conditions on the dynamics of the model, namely auto-correlation or perhaps another higher order moment dependence. In other words, we can have many models with the same invariant distribution.

The main objective in this thesis is the use of latent structures in the construction and estimation of stationary models, once an invariant distribution has been set to belong to a specific parametric family. Although some approaches are already available in the literature, most lead to models where parameter estimation can not be done using standard methods. However, a novel idea to construct stationary time series models has been introduced in Pitt et al. [78] and Pitt and Walker [79]. This idea, as we will see in Chapter 3, is based on the reversibility property and depicts the time dependence structure underlying to the model through a latent decomposition that allows for more flexible estimation methods. Therefore, most of the efforts in this thesis are devoted to further explore and generalize the approach introduced in these papers. In addition, particular interest is given to models with financial applications.

1.1 STRUCTURE OF THE THESIS

The structure of the thesis is divided in eight chapters. Chapter 1 has been confined to the present discussion, the remaining chapters are described as follows:

Chapter 2:

In order to be able to follow the results in this thesis, some basic concepts concerning stochastic processes have to be considered. This chapter serves as the background material to provide a quick reference to the reader while following subsequent chapters. Most of the models introduced in this thesis will satisfy the Markov property. In particular, the stationarity property for Markov processes is of interest. Hence, the first two sections of this chapter introduce the general definitions of stochastic, stationary and Markov processes. In particular some conceptual points required to follow Chapter 7 are also given. In section 2.3, a small discussion on diffusion processes is presented. This is mainly needed to understand the representation of some continuous-time models in Chapter 6. Finally, sections 2.4, 2.5 and 2.6 present some definitions and results of processes with independent increments, self-decomposable random variables and Ornstein-Uhlenbeck type processes required for Chapter 5.

Chapter 3:

As we have mentioned before, the construction of stationary processes with given invariant measures is a relatively new approach and has been done from different areas. In Section 3.2 we give some references. Of particular interest are the papers of Pitt et al. [78] and Pitt and Walker [79]. The idea presented in these papers was introduced for time series analysis, in this chapter we further discuss, analyze and generalize this idea to more general models. The underlying latent structure to the latter model, provides with alternative methods for model estimation. Section 3.7 of this chapter is devoted to analyze some of the possible estimation techniques available.

Chapter 4:

Since different models can have the same invariant distributions, the construction of stationary processes with given invariant distributions requires specific ways to depict the dependence structure. Pitt and Walker [79] proposed a discrete-time model where this

dependence is described through an ARCH(1)-type model with Student-t distributed innovations. In this chapter, we generalize this construction to allow for more general innovations. In particular, we give a stationary ARCH(p)-type model with generalized hyperbolic distributed innovations. The structure of this model allows us to estimate its parameters using expectation maximization methods. In Sections 4.8-4.11 we give some examples based on simulated data and real financial data.

Chapter 5:

Another innovative technique to construct continuous-time stationary models with invariant distribution belonging to the class of self-decomposable distributions is based on Ornstein-Uhlenbeck type processes, see Barndorff-Nielsen and Shephard [13]. As we will see in Chapter 2, an important component of the latter processes is given through their innovation part. In Chapter 5 we give a novel approach to describe the distribution of the innovation part corresponding to a self-decomposable random variable. In particular, this approach will help to depict the innovation part of a Ornstein-Uhlenbeck type processes. Although certain latent decomposition is also used, the material presented in this chapter separates itself from the rest of the thesis.

Chapter 6:

The availability of the data, resulting from certain phenomenon, may come in a different way of that imposed by the model. For example, stocks in financial markets are typically modelled with continuous-time processes whereas the underlying data are available discretely. The latter example constitutes an important splitting factor between the econometric and the mathematical finance theory as well as an splitting factor on the corresponding model-optimization techniques. Some approaches bridging the gap between discrete and continuous time models are available, see for example Nelson [77]. In this chapter, we extend the idea of Pitt et al. [78] to the continuous time setting. In particular, some well-known diffusion models such as Gaussian Ornstein-Uhlenbeck processes and Cox-Ingersoll-Ross models are weakly represented through a latent structure. This representation allows us to introduce a novel approach to estimate the parameters underlying to such stationary diffusion processes. Some illustrations are also given in this case.

Chapter 7:

With the idea introduced by Pitt et al. [78] certain dependence structure, given via a parametric family, has to be imposed when constructing stationary models with given invariant distributions. In this chapter we relax this assumptions by allowing such dependence structure to be given nonparametrically. In order to accomplish this task we have borrowed some ideas from the Bayesian nonparametric literature.

Chapter 8:

In this final chapter, we discuss issues highlighting some relevant points of the thesis. Some research topics to be consider in the future are also mentioned.

1.2 LIST OF SYMBOLS

We will use the following abbreviations and mathematical symbols:

Symbol	**Explanation**
$\mathbb{N}$/ $\mathbb{N}_0$	Set of positive/non-negative integer numbers
$\mathbb{R}/\mathbb{R}_+$	Set of real/non-negative real numbers
$\mathrm{Be}(\alpha, \beta)$	Beta distribution with mean $\alpha/(\alpha+\beta)$
$\mathrm{Bi}(n, p)$	Binomial distribution with mean np
$\mathrm{Ga}(\alpha, \beta)$	Gamma distribution with mean α/β
$\mathrm{GH}(\lambda, \alpha, \beta, \delta, \mu)$	Generalized hyperbolic distribution; See Section 4.2
$\mathrm{GIG}(\lambda, \delta, \gamma)$	Generalized inverse Gaussian distribution; See Section 4.2
$\mathrm{IG}(\delta, \gamma)$	Inverse Gaussian distribution with mean δ/γ
$\mathrm{N}(\mu, \sigma)$	Normal/Gaussian distribution with mean μ and variance σ
$\mathrm{Po}(\lambda)$	Poisson distribution with mean λ
$\mathrm{St}(\mu, \sigma^2, \nu)$	Non-central Student-t distribution with location parameter μ, dispersion σ and ν degrees of freedom
$\mathrm{U}(a, b)$	Uniform distribution on $[a, b]$
$\mathrm{We}(\gamma, \beta)$	Weibull distribution with density function proportional to $x^{\gamma-1}e^{-x^\gamma/\beta}$
Note:	If the random variable $X \sim \mathrm{Ga}(\alpha, \beta)$, then its (density/mass) function will be denoted by $\mathrm{Ga}(x; \alpha, \beta)$
$P(A)$	Probability of the event A
$\mathcal{P}(\cdot)$	Random probability measure
$Q \ll P$	Q is absolutely continuous with respect to P
$\mathbb{E}_Q(\cdot)$	Expectation taken under the Q measure.
$\mathbb{1}(\cdot)$	Indicator function
$\mathcal{L}_X(\cdot)$ / $\mathcal{C}_X(\cdot)$	Laplace transform/Characteristic function corresponding to X
$\overline{\mu}(\cdot)$	Tail measure corresponding to the measure $\mu(\cdot)$
$\delta_x(\cdot)$	Measure giving unit mass to the point x
$\Gamma(\cdot)$	Gamma function
$I_\nu(\cdot)$ / $K_\nu(\cdot)$	Bessel function of the first kind/third kind and order ν
$\overset{\mathrm{d}}{=}$, $\sim/\overset{\mathrm{d}}{\to}$	Equality/Convergence in distribution

CHAPTER 2

BACKGROUND AND RESULTS ON STOCHASTIC PROCESSES

This chapter gives the relevant background material on stochastic processes necessary for the subsequent chapters. The exposition presented here is devoted to definitions, results and references. For a systematic treatment of stochastic processes we refer to Gikhman and Skorokhod [40], Rogers and Williams [84, 85] and Sato [91].

2.1 STOCHASTIC PROCESSES

Definition 2.1. Let $\mathcal{T}$ be a certain index set, $(\mathrm{E}, \mathcal{E})$ a measurable space and $(\Omega, \mathcal{A}, \mathrm{P})$ a probability space. A ***stochastic processes*** is defined as a family of $(\mathrm{E}, \mathcal{E})$-valued random variables indexed on $\mathcal{T}$. We will denote it by

$$\mathrm{X} = \{X(t); t \in \mathcal{T}\} = \{X(t, \omega); t \in \mathcal{T}, \omega \in \Omega\}.$$

For fixed t, $X(t, \omega) : \Omega \to \mathrm{E}$ is $\mathcal{E}$-measurable random variable and for a fixed outcome ω, $X(t, \omega) : \mathcal{T} \to \mathrm{E}$ is a map called the ***realization*** or ***sample path*** of the stochastic process X. The alternative notation $X_t(\omega)$, will be used and the suppression of the arguments t and ω, when unnecessary, will be applied[1]. The space $(\mathrm{E}, \mathcal{E})$ is called the ***state-space*** of the process.

A common interpretation of the parameter $t \in \mathcal{T}$ is as time, however, care should

[1]Typically, when studying properties of sample paths such as continuity and differentiability, the argument 'ω' is used to stress the dependence on the probability space.

be taken in the context of nonparametric Bayesian analysis, where certain stochastic process can be used to construct random probability measures. See for example the random measures based on log-Gaussian processes used in Chapter 7. The nature of the state-space E induces a classification of the stochastic process X, for instance in the Markov process framework, when $\mathrm{E} = \mathbb{R}$, then X is referred to as a general Markov process whereas when $\mathrm{E} = \mathbb{N}$, we refer to X as a Markov chain. In a similar way when $\mathcal{T} = \mathbb{N}$, we speak of a ***discrete-time*** process and when $\mathcal{T} = \mathbb{R}$, we speak of a ***continuous-time*** process.

An alternative view of the stochastic process $\{X(t); t \in \mathcal{T}\}$ is given by its paths. That is, if $\mathrm{E}^{\mathcal{T}}$ denotes the set of all functions from $\mathcal{T}$ to E it is natural to introduce probability measures on this function space. With this, a realization of the stochastic process $\{X(t)\}$ can be seen as the outcome of a $(\mathrm{E}^{\mathcal{T}}, \mathcal{E}^{\mathcal{T}})$-random variable. Here $\mathcal{E}^{\mathcal{T}}$ denotes a suitable σ-field on $\mathrm{E}^{\mathcal{T}}$. The continuity of the paths induces a classification for X; see Definition 2.4 below.

The fundamental construction of a probability measure on function spaces is given by Kolmogorov's existence theorem. This theorem ensures the existence of such a measure when all finite-dimensional distributions are given and satisfy the consistency conditions given below.

Definition 2.2. The ***finite-dimensional distributions (fdds)*** of the stochastic process X are the distributions of the finite-dimensional vectors, that is

$$\mathrm{P}\big(X(t_1,\omega) \in A_1, \ldots, X(t_n,\omega) \in A_n\big) = \mathrm{F}_{t_1,\ldots,t_n}(A_1, \ldots, A_n), \tag{2.1}$$

where $A_i \in \mathcal{E}$, $t_i \in \mathcal{T}$, $i = 1, 2, \ldots, n$; and finite $n = 1, 2, \ldots$

Here F defines a measure on $(\mathrm{E}^n, \mathcal{E}^n)$. If F is symmetric with respect to any permutation ϱ, namely,

$$\mathrm{F}_{t_1,\ldots,t_n}(A_1, \ldots, A_n) = \mathrm{F}_{t_{\varrho 1},\ldots,t_{\varrho n}}(A_{\varrho 1}, \ldots, A_{\varrho n})$$

and

$$\mathrm{F}_{t_1,\ldots,t_{n-1},t_n}(A_1, \ldots, A_{n-1}, \mathrm{E}) = \mathrm{F}_{t_1,\ldots,t_{n-1}}(A_1, \ldots, A_{n-1}) \tag{2.2}$$

holds true for any $t_1, \ldots, t_n$ and sets $A_1, \ldots, A_n \in \mathcal{E}$, then F is said to satisfy the Kolmogorov's ***consistency conditions***.

One version of Kolmogorov's existence theorem is given as follows:

Theorem 2.1. Assume that E is a complete separable metric space. If a family of *fdds* possesses the consistency conditions (2.2), then there exists a stochastic process X such that (2.1) holds for all n and all $t_1, \ldots, t_n \in \mathcal{T}$, $A_1, \ldots, A_n \in \mathrm{E}$.

In words, Kolmogorov's existence theorem states that given any consistent family of *fdds* there is a probability space and a stochastic process defined on it whose *fdds* coincide with the given ones. See, for example, Rogers and Williams [84]. It is worth noting that within the context of this thesis the spaces $E, \mathcal{T}$ are limited to either $\mathbb{N}, \mathbb{N}_0, \mathbb{R}, \mathbb{Z}$ or their positive versions.

Example 2.1. (***Gaussian processes***) Let us assume that $\mathcal{T} = E = \mathbb{R}$. Consider the functions $\mu : \mathbb{R} \to \mathbb{R}$ and $\sigma : \mathbb{R}^2 \to \mathbb{R}$ with the latter being symmetric and positive (semi) definite on $\mathbb{R}$. That is, for any vectors $(t_1, \ldots, t_n)$ and $(c_1, \ldots, c_n)$ with $c_i \in \mathbb{R}$

$$\sum_{i=1}^{n} \sum_{j=1}^{n} c_i c_j \, \sigma(t_i, t_j) \geq 0,$$

for any choice of n. Define

$$\begin{aligned} \mathrm{F}_{t_1,\ldots,t_n}(A_1, \ldots, A_n) &= \int_{A_1} \cdots \int_{A_n} \mathrm{N}_n(\mathbf{x}; \mu, \boldsymbol{\Sigma}) \, dx_1 \cdots dx_n \\ &= \int_{A_1} \cdots \int_{A_n} |2\pi\Sigma|^{-\frac{1}{2}} \exp\left\{ -\frac{1}{2} (\mathbf{x} - \mu)' \Sigma^{-1} (\mathbf{x} - \mu) \right\} dx_1 \cdots dx_n, \end{aligned} \tag{2.3}$$

where $\mathbf{x} = (x_1, \ldots, x_n)'$ and $\mathrm{N}_n(\mathbf{x}; \mu, \boldsymbol{\Sigma})$ denotes the density of a n-dimensional normal distribution with mean vector $\mu := \{\mu(t_i); i = 1, \ldots, n\}$ and covariance matrix

$$\boldsymbol{\Sigma} := \{\sigma(t_i, t_j); i, j = 1, \ldots, n\}.$$

It can be easily verified that both, the quadratic form $(\mathbf{x} - \mu)' \Sigma^{-1} (\mathbf{x} - \mu)$ and the determinant $|\Sigma|$ are invariant under permutations of $\mathbf{x}$ and $(t_1, \ldots, t_n)$. Additionally the $(n-1)$-dimensional marginal obtained from the n-dimensional distribution (2.3) is also multivariate normal. Hence the consistency conditions (2.2) are satisfied and the

Kolmogorov's existence theorem implies the existence of a stochastic process known as a Gaussian process on $\mathbb{R}$. See Abrahamsen [1]

Let $\mathcal{P}$ denote the measure on $(\mathrm{E}^{\mathcal{T}}, \mathcal{E}^{\mathcal{T}})$ constructed via Theorem 2.1. Even though, this theorem provides information about the existence of a measure $\mathcal{P}$ on, perhaps, the infinite dimensional space $(\mathrm{E}^{\mathcal{T}}, \mathcal{E}^{\mathcal{T}})$, it is not so straightforward to deduce some properties of the sample paths, such as continuity. Therefore, for continuous-time processes further properties must be added.

Continuity of stochastic processes or even more general random functions is a property related with the convergence of random sequences $\{X(t_n)\}$. As there are different types of convergence for random variables there are different types of continuity for stochastic processes. Before introducing these types of continuity the concept of separability is defined.

Definition 2.3. A stochastic process $\mathrm{X} = \{X(t); t \in \mathcal{T}\}$ with state-space $(\mathrm{E}, \mathcal{E})$ is said to be ***separable*** if there exist a countable set $S \subset \mathcal{T}$ such that for every set $I \subset \mathcal{T}$ and every closed set $K \in \mathrm{E}$, the sets

$$A = \{\omega; X(t, \omega) \in K, t \in I\}$$

and

$$B = \{\omega; X(t, \omega) \in K, t \in S \cap I\}$$

differ by a set Ξ such that $\mathrm{P}[\Xi] = 0$.

Separability ensures that the *fdds* characterize certain sample path properties by requiring that sample paths are determined by their values on an everywhere dense but countable subset of positions in E.

Definition 2.4. (Continuity of stochastic processes)
Let $\mathrm{X} = \{X(t); t \in \mathcal{T}\}$ be a continuous-time stochastic process:

- (i) X is said to be ***continuous in probability*** or ***stochastically continuous*** if for all $\epsilon > 0$ and all $t \in \mathcal{T}$

$$\mathrm{P}\left(|X_{t+h} - X_t| \geq \epsilon\right) \to 0 \text{ as } h \to 0.$$

- (ii) X is said to be ***continuous in p-th mean*** if for all $t \in \mathcal{T}$

$$\mathbb{E}[|X_{t+h} - X_t|^p] \to 0 \text{ as } h \to 0.$$

- (iii) X is said to be ***almost surely continuous*** if for all $t \in \mathcal{T}$

$$\mathrm{P}\left(\lim_{h\to 0} |X_{t+h} - X_t| = 0\right) = 1.$$

- (iv) A separable stochastic process X is said to be ***almost surely path continuous*** if

$$\mathrm{P}\left(\bigcup_{t\in\mathcal{T}} \lim_{h\to 0} |X_{t+h} - X_t| \neq 0\right) = 0.$$

A method to clarify some of the continuity properties of a process X, is given by the existence of a ***modification***, that is, a process $\tilde{\mathrm{X}}$ such that $\mathrm{P}[\omega : X(t,\omega) = \tilde{X}(t,\omega)] = 1$ for any $t \in \mathcal{T}$. It is possible to see that for each well-defined stochastic process a separable modification exists, see Doob [27]. Even though it is difficult to check property (iv) for a given stochastic process, it is sometimes possible to work with a modification of it. The following theorem gives a sufficient condition for a stochastic process to have a continuous modification.

Theorem 2.2. (Kolmogorov's continuity criterion)
Let $\{X(t); t \in \mathcal{T}\}$ be a stochastic process on $\mathcal{T} = [0,T]$ with complete separable state space $(\mathrm{E}, \mathcal{E})$. A continuous modification $\tilde{\mathrm{X}}$ of X exists if there are numbers $K > 0, \alpha > 0, \beta > 0$, such that

$$\mathbb{E}\left[|X_t - X_s|^\alpha\right] \leq K|t-s|^{1+\beta}$$

for $0 \leq s, t \leq T$.

For a proof of this theorem see Gikhmam and Skorokhod [40].

A classification of stochastic processes that will play an important role in this thesis is given by the following definition.

Definition 2.5. A stochastic process X is said to be ***strictly stationary*** if for arbitrary h, n and $t_1, t_2, \ldots, t_n$, with $t_i, t_i + h \in \mathcal{T}$, $i = 1, 2, \ldots, n$, its *fdds* are invariant under shifts of time, that is,

$$\{X_{t_1}, X_{t_2}, \ldots, X_{t_n}\} \sim \{X_{t_1+h}, X_{t_2+h}, \ldots, X_{t_n+h}\},$$

where $\sim$ stands for the identity of the distributions.

A Gaussian process with $\mathcal{T} = \mathbb{N}$ or $\mathcal{T} = \mathbb{R}$ is strictly stationary if its mean function is constant, $\mu(t) = a$, and its covariance function depends on the difference $(s - t)$ only, that is, $\sigma(s, t) = \sigma(0, s - t)$. Strictly speaking, the latter conditions define a covariance stationary process. However, in the case of Gaussian processes, that are completely characterized by the mean and covariance functions, covariance stationarity implies strictly stationarity. See Ibragimov and Rozanov [43].

2.2 MARKOV PROCESSES

One of the most widely studied families of stochastic processes is the family of Markov processes. An easy way of understanding the main features of such processes is given when the set $\mathcal{T}$ is understood as time. Then, under this appreciation, a Markov process is a stochastic process whose future evolution in time only depend on the present. Before giving the definition of Markov processes, let us introduce some concepts.

A probability space $(\Omega, \mathcal{F}, \mathrm{P})$ with a filtration $(\mathcal{F}_t)_{t\geq 0}$ is often called a ***filtered probability space*** and denoted by $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t\geq 0}, \mathrm{P})$. It is commonly assumed that the σ-algebra $\mathcal{F}_0$ contains all the P-null sets of $\mathcal{F}$ and that the filtration $(\mathcal{F}_t)_{t\geq 0}$ is right continuous.

For a filtration we mean an increasing family of σ-algebras, that is $\mathcal{F}_s \subset \mathcal{F}_t$ if $s < t$, and we say that it is right continuous if $\mathcal{F}_t = \bigcap_{u>t} \mathcal{F}_u$ for all t, $0 \leq t \leq \infty$. A Filtration is usually understood as a stream of information resulting from a stochastic process.

Definition 2.6. Let $\mathrm{X} = \{X(t); t \in \mathcal{T}\}$, $\mathcal{T} = \mathbb{N}$ or $\mathcal{T} = \mathbb{R}_+$ be a stochastic process defined on a filtered probability space $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t\geq 0}, \mathrm{P})$. This process is said to be **Markov** with state-space E, if X is adapted to the filtration $(\mathcal{F}_t)_{t\geq 0}$ and for every $\mathcal{E}$-measurable bounded function h we have

$$\mathbb{E}\left[h(X_u) \mid \mathcal{F}_t\right] = \mathbb{E}\left[h(X_u) \mid \sigma(X_t)\right], \tag{2.4}$$

where $u \geq t$ and $\sigma(X)$ denotes the σ-algebra generated by X.

In particular if $h(x) = e^{-x\xi}$, with $\xi \in \mathbb{R}_+$, then from (2.4) the agreement of Laplace transforms follows. An important result in the theory of Laplace transforms says that if (2.4) holds for the latter function then it holds for every h of the form stated in Definition 2.6. One could think of the property (2.4) as the fact that the best prediction of the future given the past and the present depends only on the present. This property is known as the ***Markov property***.

A function $\mathrm{P}_{s,t}(x, A) = \mathrm{P}(X_t \in A \mid X_s = x)$ on $\mathcal{T} \times \mathrm{E} \times \mathcal{E}$, $t > s$, $x \in E$, with values in $[0, 1]$ that is $\mathcal{E}$-measurable for every $A \in \mathcal{E}$ and satisfies $\mathrm{P}_{s,t}(x, \mathrm{E}) = 1$ is called a ***Markov transition function*** [2] if

$$\mathrm{P}_{s,u}(x, A) = \int_{\mathrm{E}} \mathrm{P}_{s,t}(x, dy)\, \mathrm{P}_{t,u}(y, A), \qquad s < t < u, \quad x \in \mathrm{E}, \quad A \in \mathcal{E}. \tag{2.5}$$

Relation (2.5) is known as ***Chapman-Kolmogorov equation***. If $\mathrm{P}_{s,t}$ depends only on the difference $t - s$, that is, $\mathrm{P}_{s,t} = \mathrm{P}_{0,t-s} = \mathrm{P}_{t-s}$, then the transition function is said to be ***time homogeneous*** [3]. In a similar way a transition function $\mathrm{P}_{s,t}(x, A)$ is said to be ***spatially homogeneous*** (or translation invariant) if $\mathrm{P}_{s,t}(x, A) = \mathrm{P}_{s,t}(0, A - x)$, where $A - x = \{y - x; y \in A\}$.

It can be seen that a stochastic process X with values in $(\mathrm{E}, \mathcal{E})$ and adapted to the filtration $(\mathcal{F}_t)_{t\geq 0}$ is a (homogeneous) Markov process with transition function $\mathrm{P}_{s,t}(x, A)$ $(\mathrm{P}_t(x, A))$ if for each $0 \leq s < t$, $A \in \mathcal{E}$

$$\mathrm{P}(X_t \in A \mid \mathcal{F}_s) = \mathrm{P}_{s,t}(X_s, A) \qquad (\mathrm{P}_{t-s}(X_s, A), \text{ resp}).$$

[2] In the case where the Markov process X exploits or dies (or even in the general case) it is sometimes convenient to consider transition functions with $\mathrm{P}_{s,t}(x, \mathrm{E}) \leq 1$ (sub-Markov). A process with this feature can always be extended from sub-Markov to Markov just by adding a "cementery" state. See Rogers & Williams [84].

[3] Homogeneous transition functions are alternatively named as stationary transition functions.

The Markov property (2.4) can also be stated in terms of the transition functions, as follows,

$$\mathbb{E}\left[h(X_t) \mid \mathcal{F}_s\right] = \mathrm{P}_{s,t}(X_s, h) \qquad (\mathrm{P}_{t-s}(X_s, h), \text{ resp}),$$

where

$$\mathrm{P}_{s,t}(x, h) = \mathrm{P}_{s,t}h(x) = \int_{\mathrm{E}} h(y)\,\mathrm{P}_{s,t}(x, dy),$$

for h bounded and $\mathcal{E}$-measurable. In the above notation $\mathrm{P}_{s,t}$ denotes a family of positive bounded operators. Using this notation, the Chapman-Kolmogorov equation becomes the semigroup property $\mathrm{P}_{s,t+u} = \mathrm{P}_{s,t}\mathrm{P}_{s,u}$.

A Markov transition probability is said to be ***standard*** or ***normal*** if

$$\lim_{t \downarrow 0} \mathrm{P}_t(x, \cdot) = \delta_x(\cdot),$$

for all $x \in \mathrm{E}$. All transition probabilities to be consider will be assumed to be standard.

Using the Chapman-Kolmogorov equation (2.5) we have that for a standard transition probability

$$\mathrm{P}_{t+\varepsilon}(x, A) = \int_{\mathrm{E}} \mathrm{P}_t(x, dy)\,\mathrm{P}_\varepsilon(y, A), \quad \varepsilon > 0,$$

and consequently

$$\lim_{\varepsilon \downarrow 0} \mathrm{P}_{t+\varepsilon}(x, A) = \mathrm{P}_t(x, A),$$

from where the right continuity of the transition probabilities follows.

Sometimes there is need to consider a property, stronger than the Markov property, which allows the time of a process to be random. A property with this feature is the strong Markov property which roughly speaking establishes that, the Markov property is satisfied if the times t are replaced with certain random times, independent of the future, called stopping times.

Definition 2.7. The process $\mathrm{X} = \{X(t); t \in \mathcal{T}\}$ is said to be **strong Markov** if X is adapted to the filtration $(\mathcal{F}_t)_{t \geq 0}$ and satisfies the following:

a) For each stopping time T, $t \in \mathcal{T}$ and $A \in \mathcal{E}$ the random variable
$\mathbb{1}(T < \infty)\,\mathrm{P}_{T,T+t}\,(X_T, A)$ is $\mathcal{F}_t$-measurable.

b) For every stopping time T, $t \in \mathcal{T}$ and $A \in \mathcal{E}$,

$$\mathrm{P}\left(\{X_{T+t} \in A\} \cap \{T < \infty\} \mid \mathcal{F}_T\right) = \mathrm{P}_{T,T+t}\left(X_T, A\right) \mathbb{1}(T < \infty) \quad a.s. \tag{2.6}$$

Equality (2.6) is known as the ***Strong Markov property***. When the process is time homogeneous, condition a) is not needed, since the right hand side of (2.6) will be always $\mathcal{F}_T$-measurable.

Proposition 2.1. (Tudor [98])
If we put $\mathcal{T} = \mathbb{N}_0$, then each Markov process is a strong Markov process.

Proof. Let T be a stopping time. Using the equality $\{T < \infty\} = \cup_{s=0}^{\infty}\{T = s\}$ it is enough to prove that

$$\mathrm{P}\left(\{X_{T+t} \in A\} \cap B \cap \{T = s\}\right) = \int_{B\cap\{T=s\}} \mathrm{P}_{T,T+t}(X_s, A)\, d\mathrm{P}, \tag{2.7}$$

for each $A \in \mathcal{E}$, $B \in \mathcal{F}_T$ and $s \in \mathcal{T}$. Equality (2.7) can be written as follows

$$\mathrm{P}\left(\{X_{s+t} \in A\} \cap B \cap \{T = s\}\right) = \int_{B\cap\{T=s\}} \mathrm{P}_{s,s+t}(X_T, A)\, d\mathrm{P}. \tag{2.8}$$

Since $B \cap \{T = s\} \in \mathcal{F}_s$, then (2.8) is a immediate consequence of the Markov property and the result follows. □

In other words, Proposition 2.1 says that any discrete-time Markov process is also a strong Markov process.

The following definition will be important in Chapter 6, when relating certain stochastic processes to diffusion processes.

Definition 2.8. (Karlin and Taylor [53]) A Markov process $\mathrm{X} = \{X(t); t \geq 0\}$ without killing is said to be a ***standard process*** if it has ***càdlàg***[4] paths and satisfies the quasi-left continuity condition[5].

It can be proved that every stochastically continuous strong Markov process has a modification which is a standard process. See Rogers and Williams [84].

[4] Càdlàg is the acronym in French for "continu à droite et pourvu de limites à gauche", right continuous functions with left limits.

[5] For a definition of quasi-left continuity see Karlin and Taylor [53] or Rogers and Williams [84].

All the *fdds* of a Markov process X with state-space $(\mathrm{E}, \mathcal{E})$ are expressible in terms of its ***initial measure*** $\mu_0(A) = \mathrm{P}(X_0 \in A)$ and its transition probability $\mathrm{P}_{s,t}(x, A)$ as follows

$$\begin{aligned} & \mathbb{E}\left[h(X_{t_1}, X_{t_2}, \ldots, X_{t_n})\right] \\ = & \int \mu_0(dx_0) \int \mathrm{P}_{0,t_1}(x_0, dx_1) \int \ldots \int \mathrm{P}_{t_{n-1},t_n}(x_{n-1}, dx_n)\, h(x_1, x_2, \ldots, x_n), \end{aligned} \tag{2.9}$$

where h is $\mathcal{E}^n$-measurable.

In the time homogeneous case, a measure μ satisfying

$$\mu(A) = \int_{\mathrm{E}} \mathrm{P}_t(x, A)\, \mu(dx). \tag{2.10}$$

for any time t and $A \in \mathcal{E}$, is called an ***invariant measure*** or ***stationary distribution*** when $\mu(\mathrm{E}) = 1$. Using equations (2.9) and (2.10) it is clear that any time homogeneous Markov process with invariant probability measure μ is a stationary process, since any shifted version of the *fdds* preserves the same invariant/initial probability measure. Most of the processes to be analyzed in this thesis are time homogeneous Markov processes, in which case the existence of stationary distributions translates to valid stationary processes.

A probability measure μ is said to be the ***limit*** or ***ergodic distribution*** of a time-homogeneous Markov process X if $\mathrm{P}_t(x, \cdot) \to \mu(\cdot)$ as $t \to \infty$, for any $x \in \mathrm{E}$.

Example 2.2. Let $\mathrm{X} = \{X(t); t \in \mathcal{T}\}$ be a Gaussian process with mean and covariance function μ and σ respectively. If

$$\sigma(s, u) = \frac{\sigma(s, t)\, \sigma(t, u)}{\sigma(t, t)}, \quad s \leq t \leq u, \quad s, t, u \in \mathcal{T}, \tag{2.11}$$

then the Gaussian process is Markov. Furthermore, if $\mathcal{T} = \mathbb{R}$ and $\sigma(s, t) = ae^{-b|s-t|}$ then X is stationary. A zero mean stationary Gaussian Markov process is called a ***Gaussian Ornstein-Uhlenbeck process***. For a proof of these statements, see Kallenberg [52]. ∘

2.3 DIFFUSION PROCESSES

A continuous-time (strong) Markov process with (almost surely) continuous sample paths is referred to as a ***diffusion process***.

There are two ways of defining diffusion processes; the starting points are either the Markov transition probabilities (2.5), satisfying some regularity conditions, or the evolution in time of the random variables $\{X(t); t \in \mathcal{T}\}$, in other words the sample path description of the process. The latter is typically given through a stochastic differential equation (SDE). A connection between the two approaches can be given by assuming that the process with given Markov transition probabilities, used in the former approach, is standard. This connection is by no means trivial, we refer to Rogers and Williams [84] for an exhaustive treatment on this issue.

In order to depict a Diffusion processes we need the definition of a Brownian motion.

Definition 2.9. A continuous-time stochastic process $\mathrm{W} = \{W_t; t \geq 0\}$ is said to be a ***Brownian motion*** if

- $W_0 = 0$
- $\mathrm{W} = \{W_t; t \geq 0\}$ is a Gaussian process with mean 0 and $\mathbb{E}[W_t\, W_s] = \min\{t, s\}$.

Other definitions of Brownian motion are available, for example, as an independent increments process (introduced below) with the law of the increments being normally distributed. In particular, Definition 2.9 implies that the process W is stochastically continuous with stationary and independent increments.

Typically, one-dimensional diffusion processes can be introduced as the solution of a stochastic differential equation, which can be symbolically written as

$$dX_t = \mu(X_t)\, dt + \sigma(X_t)\, dW_t, \quad X_0 = x \tag{2.12}$$

where $\mathrm{W} = \{W_t; t \geq 0\}$ is a Brownian motion and $\mu(\cdot)$ and $\sigma(\cdot)$ are given functions termed the ***drift*** and ***diffusion*** coefficient respectively. In the above equation dW_t denotes the "differential" of a Brownian motion, which is not taken in the classical sense. More formally, we can interpret equation (2.12) as the stochastic integral

equation

$$X_t = X_0 + \int_0^t a(X_s)\,ds + \int_0^t \sigma(X_s)\,dW_s. \tag{2.13}$$

Where the first integral is a Riemann integral and the second integral is defined as

$$\int_0^t \sigma(X_s)\,dW_s = \lim_{n\to\infty} \sum_{j=0}^{\lfloor 2^n t\rfloor - 1} \sigma\left(X_{\frac{j}{2^n}}\right)\left(W_{\frac{j+1}{2^n}} - W_{\frac{j}{2^n}}\right), \tag{2.14}$$

where $\lfloor a \rfloor$ denotes the largest integer no larger than a. In general, the above integral is known as the Itô integral. A process defined through (2.13) is known as an Itô process; see Rogers and Williams [84, 85]. Sometimes it is also of interest to consider measurable transformations of diffusion processes. The following result, known as the Itô lemma, generalizes the chain rule of ordinary calculus.

Theorem 2.3. (Itô lemma)
Suppose that $g(x,t)$ is a twice continuous differentiable function on $[0,T] \times \mathbb{R}$. Then for any Itô process X, the process defined through the transformation $Y_t = g(X_t,t), t \in [0,T]$, follows an Itô process. Furthermore, its canonical decomposition is given by the formula

$$dY_t = g_t(X_t,t)\,dt + g_x(X_t,t)\,a(X_t)\,dt + g_x(X_t,t)\,\sigma(X_t)\,dW_t + \frac{1}{2}g_{xx}(X_t,t)\,\sigma_t^2\,dt, \tag{2.15}$$

where $g_t = \partial g/\partial t$, $g_x = \partial g/\partial x$ and $g_{xx} = \partial^2 g/\partial x \partial x$.

In Chapter 6, we will focus on conditions of a standard Markov process X with transition probabilities $\mathrm{P}_t(x,\cdot)$ that allow us to relate it (in a weak sense) with a diffusion process. We will mainly concentrate on ***regular diffusions***, that is, processes with the property that starting in the interior of the state space, there exists a positive probability to reach any other state in such interior set.

Theorem 2.4. (Karlin and Taylor [53])
Let $\mathrm{X} = \{X(t); t \geq 0\}$ be a standard Markov process and suppose that

$$\lim_{h \downarrow 0} \frac{1}{h}\mathrm{P}\left(|X_{t+h} - X_t| > \varepsilon \mid X_t = x\right) = 0 \tag{2.16}$$

is satisfied. Then X is a diffusion process.

Condition (2.16) is known as the ***Dynkin condition***.

An application of the Chebyshev inequality ensures that (2.16) is satisfied if

$$\lim_{h\downarrow 0}\frac{1}{h}\mathbb{E}[|X(t+h)-X(t)|^p \mid X(t)=x]=0, \qquad \text{for } p>2, \tag{2.17}$$

for all x in the interior of E.

If the above condition is satisfied the drift and diffusion coefficients corresponding to the diffusion process are given through

$$\mu(x,t) = \lim_{h\downarrow 0}\frac{1}{h}\mathbb{E}[X(t+h)-X(t) \mid X(t)=x] \tag{2.18}$$

$$\sigma^2(x,t) = \lim_{h\downarrow 0}\frac{1}{h}\mathbb{E}[(X(t+h)-X(t))^2 \mid X(t)=x]. \tag{2.19}$$

Hence, if we are given a model with transition probabilities that satisfy the Dynkin condition, then we can associate to it a diffusion process given as the weak solution of a SDE with certain boundary conditions.

On the other hand, if we are given an equation such as (2.12), then it can be proved that sufficient conditions for the existence and uniqueness of a strong solution are given by ensuring that the coefficients μ and σ are ***Lipschitz continuous*** in the following sense

$$\begin{aligned} |\mu(x)-\mu(y)|+|\sigma(x)-\sigma(y)| &\leq C|x-y| \\ |\sigma(0)|+|\mu(0)| &\leq C, \end{aligned} \tag{2.20}$$

where $C>0$, $\mu(\cdot,\cdot):[0,T]\times \mathrm{E}\to \mathrm{E}$, $\sigma(\cdot,\cdot):[0,T]\times \mathrm{E}\to \mathrm{E}\times\mathrm{E}$ and $x,y\in \mathrm{E}$. Condition (2.20) can be generalized to allow explosion in the coefficients as time increases. This, can be done by replacing C in (2.20) by a finite-valued increasing function C_t. In addition ***linear growth*** conditions may be added to the coefficients to avoid their explosion. See Rogers and Williams [85].

The advantage of having strong solutions to equation (2.12) is that it is possible to consider modifications with almost surely continuous sample paths whereas with weak solutions we can only consider stochastically continuous processes. See Definition (2.4).

However, within the framework of this thesis we will be interested in the law properties of a given process rather than in its path properties.

The discussion below follows Karlin and Taylor [53].

In Chapter 6, we will be interested in the stationary distribution (when available) corresponding to a given diffusion process. When the stochastic process at issue is given through a SDE such as (2.12), stationarity conditions can be analyzed with the ***scale measure*** $S(\cdot)$ and ***speed measure*** $M(\cdot)$ corresponding to the given SDE. These measures have respective Lebesgue densities given by

$$s(x) = \exp\left\{\int_{x^*}^{x} \frac{2\mu(y)}{\sigma^2(y)} dy\right\}$$

$$m(x) = (s(x)\sigma^2(x))^{-1},$$

for all x, x^* in the interior of E and x^* being arbitrary, but fixed.

Let the state-space E be of the form $(l,r), (l,r], [l,r)$ or $[l,r]$ with $-\infty \leq l < r \leq \infty$ and define

$$\psi(x) = m(x)\,[C_1 S(x^*, x) + C_2], \tag{2.21}$$

where C_1 and C_2 are constants. If it is possible to find values for the latter constants, such that: $\psi(x) \geq 0$ on E and $\int_l^r \psi(x)\,dx = 1$ are satisfied, then a stationary density exists and we denote it as $q_X(x) = \psi(x)$.

When analyzing diffusion processes one of the most important issues to consider is the behavior of such processes at their boundary points. For example, if both boundaries, l and r are entrance[6], then the underlying diffusion process has a unique stationary distribution. In particular, from equation (2.21), $C_1 = 0$ and $C_2 = M(\mathrm{E}) < \infty$ so

$$q_X(x) = \frac{m(x)}{M(\mathrm{E})}.$$

However, for other boundary classifications, results that allow us to tell if a process is stationary are not available in such generality. For instance, in both the Brownian

[6]An entrance boundary cannot be reached starting from an interior point, but is possible for the underlying diffusion to start there. A way to ensure that a boundary, say l, is entrance is checking that $S((l,x]) = \infty$ and $\int_l^{x^*} S([\eta, x^*])\,M(d\eta) < \infty$, where x^* is in the interior of E.

motion and the Gaussian Ornstein-Uhlenbeck process both boundaries are ***natural***[7]. However, whereas the Ornstein-Uhlenbeck process has a stationary distribution the Brownian motion does not. For an exhaustive treatment of boundary classifications and further results concerning stationairty, we refer to Karlin and Taylor [53]. It is worth mentioning here that the construction of stochastic processes presented in Chapter 6 describes the dynamics of the model at the interior of the corresponding state-spaces. Hence, the behavior of the processes at the boundary points has to be specified separately.

2.4 INDEPENDENT INCREMENTS PROCESSES

Processes with independent increments or additive processes play an important role in many areas of research, such as, finance and Bayesian nonparametric analysis. These processes belong to the class of Markov processes and have as a particular case the class of Lévy processes[8].

Definition 2.10. A stochastic process $\mathrm{X} = \{X(t); t \in \mathcal{T}\}$ on $\mathbb{R}$ is called an ***independent increments process*** (IIP) if it is stochastically continuous with càdlàg paths and for any $n \in \mathbb{N}$ and $0 \leq t_1 < t_2 \cdots < t_n \in \mathcal{T}$ the random variables (increments)

$$X(t_{j+1}) - X(t_j), \quad 1 \leq j \leq n$$

and $X(t_1)$ are independent. If additionally, $X(0) = 0$ and the above increments are time-homogeneous (stationary increments), then X is called ***Lévy process***. If the càdlàg requirement is dropped then we speak of a process in law. We denote IIP^0 as an independent increments process starting at zero almost surely.

The càdlàg paths condition ensures that only jump discontinuities may occur, and we say that X has a ***fixed jump*** at some time $t > 0$ if $\mathrm{P}(X_t \neq X_{t-}) > 0$. If the stationarity assumption is added to the increments of X then the possibility of fixed jumps is excluded.

Independent increments processes are Markov processes with spatially homogeneous

[7] A diffusion process can neither be reached in finite mean time nor be started from a natural boundary.

[8] Lévy [64] introduced the general structure for additive processes, for this reason some authors refer to additive processes as Lévy processes. See Sato [91] for a discussion on this issue.

transition probabilities. See Sato [91]. As we mentioned before the distributional features of a Markov process can be mainly studied via its transition probabilities. Since in the case of IIP transition probabilities are spatially homogeneous, then their distributional characteristics are equally studied through the law of the increments. Although in general a closed form for the transition probabilities of independent increments processes does not exist, it is possible to work with their characteristic function.

Theorem 2.5. (***Lévy-Khintchine representation***)
The characteristic function of a stochastically continuous independent increments process $\mathrm{X} = \{X(t); t \in \mathcal{T}\}$ on $\mathbb{R}$ is given by

$$\mathbb{E}\left[e^{i\lambda X_t}\right] = \mathbb{E}\left[e^{i\lambda X_0}\right] \exp\left\{-\psi_t(\lambda)\right\},$$

where ψ_t is known as the ***characteristic exponent*** and is given by

$$\psi_t(\lambda) = i\lambda d_t + \frac{\lambda^2}{2} G_t + \int_{\mathbb{R}} \left\{e^{i\lambda x} - 1 - h(x)\right\} N_t(dx). \tag{2.22}$$

Here, d_t is a continuous function on $\mathbb{R}$ named the drift term, $G_t \in \mathbb{R}_+$ is a continuous nonnegative symmetric linear operator with $G_{t_2} - G_{t_1}$ nonnegative for $t_1 < t_2$ named the Gaussian term. N_t is a positive measure satisfying $\int_{\mathbb{R}}(1 \wedge x^2)N_t(dx) < \infty$ called the ***Lévy measure*** and h is a bounded "truncation" function with compact support which "behaves like x" near the origin. This function typically is replaced by the canonical truncation function $H(x) = x\,\mathbb{I}(|x| \leq 1)$. See Jacod and Shiryaev [45].

An IIP with these characteristics will be referred collectively by the triplet $(d_t, G_t, N_t)_h$ or $(d_t, G_t, N_t)_H$ when the above canonical truncation is used. If an independent increments process starting at zero, $X_0 = 0$ a.s., has the characteristics $(d_t, G_t, N_t)_h = (td, tG, tN)_H$ then we have a Lévy process. When working with Lévy processes we have $\psi_t = t\psi$, where ψ is given by (2.22) by removing the dependence on t. In this case ψ will be referred to as the characteristic exponent of a Lévy process.

Independent increments processes starting at zero are closely related to infinitely divisible random variables. A random variable X is called ***infinitely divisible*** if and only if for each integer n there exists i.i.d. random variables $X_{1n}, \ldots, X_{nn}$ such that $X \stackrel{\mathrm{d}}{=} X_{1n} + \cdots + X_{nn}$. See Sato [91]. Infinitely divisible distributions are characterized by their characteristic function, which is given by $e^{-\psi(\lambda)}$, with $\psi(\lambda)$ given by (2.22).

Proposition 2.2. (Theorem 9.1 Sato [91])
If X = $\{X(t); t \in \mathcal{T}\}$ is an IIP^0, then for every t, the distribution of X_t is infinitely divisible.

The above proposition establishes an equivalence in law between infinitely divisible measures and IIP processes starting at zero. As we mentioned before equivalence in law (agreement of the *fdds*) does not necessarily guarantee the existence of a corresponding stochastic process with càdlàg paths. However, in this case the following proposition establishes that any IIP process in law has a modification which is an IIP process.

Proposition 2.3. (Theorem 11.5 Sato [91])
If $\{X(t); t \in \mathcal{T}\}$ is an IIP^0 in law, then it has a modification which is an IIP^0 with càdlàg paths.

In Chapter 5 we will consider increasing IIP processes defined on $\mathbb{R}_+$. In such cases, it is more convenient to work with the Laplace transform than with the characteristic function. The Laplace transform of increasing IIP processes is given by

$$\mathbb{E}\left[e^{-\lambda X_t}\right] = \mathbb{E}\left[e^{-\lambda X_0}\right] \exp\left\{-d_t\lambda - \int_0^\infty (1 - e^{-\lambda x})\, N_t(dx)\right\}, \tag{2.23}$$

where $\int_{\mathbb{R}_+} (1 \wedge x)\, N_t(dx) < \infty$.

For non-negative infinitely divisible distributions the Lévy-Khintchine representation simplifies to equation (2.23) with $X_0 = 0$ and the parameter d in the corresponding triplet can be understood as the left-extremity of the infinitely divisible distribution.

2.5 SELF-DECOMPOSABILITY

Definition 2.11. A random variable X defined on $\mathbb{R}^d$, is said to be ***self-decomposable***[9] (SD) on $\mathbb{R}^d$ if there exists a random variable W_ρ on $\mathbb{R}^d$ such that

$$X \stackrel{\mathrm{d}}{=} \rho X + W_\rho$$

for all $0 < \rho < 1$. The random variable W_ρ will be referred to as the corresponding ***innovation variable***.

[9] SD random variables are also known as random variables with distributions in the class L. Although, they are defined differently, Sato [91] has shown that a random variable has a distribution of class L if and only if the law of this random variable is self-decomposable.

If we denote by $C_Z(\lambda)$ the characteristic function of the random variable Z, then for a SD random variable X, the corresponding characteristic function is given by

$$C_X(\lambda) = C_X(\rho\lambda)\, C_{W_\rho}(\lambda).$$

It can be proved that the class of self decomposable random variables belongs to the class of infinitely divisible random variables. In particular, the characteristic function corresponding to SD random variables on $\mathbb{R}$ is given by $e^{-\psi(\lambda)}$ where ψ is given by (2.22) with the triplet (d, G, N), where

$$N(dx) = \frac{k(x)}{|x|}dx,$$

and $k(x)$ is a positive function which is increasing on $(-\infty, 0)$ and decreasing on $(0, \infty)$. See Wolfe [106] and Sato [91]. With this, a corresponding modification of an IIP starting at zero is available, in which case we refer to a ***self decomposable processes***. The class of SD random variables also has the property that the corresponding densities are unimodal, see Wolfe [105] and Yamazato [110].

Self decomposable random variables are related to broad-sense self-similar processes. A stochastic process $\mathrm{X} = \{X_t; t \geq 0\}$ on $\mathbb{R}^d$ is called ***broad-sense self-similar process*** if for any $a > 0$, there are $b > 0$ and a function $c(t)$ from $[0, \infty)$ to $\mathbb{R}^d$ such that

$$\{X_{at}; t \geq 0\} \stackrel{\mathrm{d}}{=} \{bX_t + c(t); t \geq 0\}.$$

If $c(t) = 0$, then the process X is called ***self-similar process*** .

Theorem 2.6. (Sato [90]) If $\mathrm{X} = \{X_t; t \geq 0\}$ is a broad-sense self-similar process on $\mathbb{R}^d$, then for every $t \geq 0$, the distribution of X_t is self-decomposable.

2.6 ORNSTEIN-UHLENBECK TYPE PROCESSES

Ornstein-Uhlenbeck processes were first studied by Wolfe [107] as a generalization of the first order auto-regressive model given by

$$X_{n+1} = \rho X_n + W_{n+1,\rho}, \tag{2.24}$$

where $|\rho| < 1$ and $\{W_{n+1,\rho}\}$ are i.i.d. innovation random variables. In particular if $|\rho| < 1$ and $\{W_{n+1,\rho}\}$ is an i.i.d sequence of random variables with zero mean and constant variance, the model (2.24) is strictly stationary, meaning that $X_{n+1} \stackrel{\mathrm{d}}{=} X_n$. Notice the similarity of (2.24) with SD random variables.

The continuous time analogue to equation (2.24) is called the ***Ornstein-Uhlenbeck type*** (OU-type) process given by

$$X_t = \rho^t X_0 + \int_0^t \rho^{t-s}\, dL_s, \tag{2.25}$$

where $\rho = e^{-\lambda}$, $\lambda > 0$ and $\mathrm{L} = \{L_t; t \geq 0\}$ is a Lévy process on $\mathbb{R}^d$. A necessary and sufficient condition for the existence of (2.25) is given in Sato [91]. See also Wolfe [107]. In the one-dimensional case this sufficiency condition is given by $\mathbb{E}[\log(1+|L(1)|)] < \infty$, or in terms of the Lévy measure of L, given by

$$\int_{|x|>0} \log|x|\, N(dx) < \infty. \tag{2.26}$$

It can be proved, by a applying Itô's formula, that expression (2.25) is the solution to the following stochastic differential equation

$$dX_t = -\lambda X_t\, dt + dL_t, \tag{2.27}$$

which is the continuous version of the difference equation corresponding to (2.24) given by

$$X_{n+1} - X_n = (\rho - 1)X_n + W_{n+1,\rho}.$$

When the Lévy process L_t in (2.27) is characterized by the triplet $(0, t, 0)_H$, that is the Brownian motion, then $\{X_t; t \geq 0\}$ is the Gaussian OU process. See Cox and Miller [23]. Also when working with OU-type processes it can be proved that a càdlàg modification exists, see Sato [90]. Another important property of these processes is that they are Markov.

OU-type processes are related with SD distributions in the sense that SD distributions can be seen as the limit distributions of OU-type processes. Therefore OU-type processes constitute a class of stochastic process with a wide class of limiting distributions. An excellent compendium of results and applications of these models is given

by Barndorff-Nielsen, Shephard and other co-workers, see [11], [10] and [12] and the references therein for a more extended treatment.

The relation between OU-type processes and SD random variables is given by the following theorem

Theorem 2.7. A random variable X is SD if and only if has a representation of the form

$$X \stackrel{\mathrm{d}}{=} \int_0^\infty e^{-\lambda s}\, dL_{\lambda s} \tag{2.28}$$

where $\lambda > 0$ and $\mathrm{L} = \{L_t, t \geq 0\}$ is a Lévy process.

From (2.28),

$$\begin{aligned} X &\stackrel{\mathrm{d}}{=} \int_{(t,\infty]} e^{-\lambda s}\, dL_{\lambda s} + \int_{(0,t]} e^{-\lambda s}\, dL_{\lambda s} \\ &\stackrel{\mathrm{d}}{=} e^{-\lambda t} \int_0^\infty e^{-\lambda s}\, dL_{\lambda(t+s)} + \int_{(0,t]} e^{-\lambda s}\, dL_{\lambda s}. \end{aligned} \tag{2.29}$$

Following the stationarity and independence of the underlying Lévy process increments, the two integrals in (2.29) are independent and

$$\int_0^\infty e^{-\lambda s}\, dL_{\lambda(t+s)} \stackrel{\mathrm{d}}{=} \int_0^\infty e^{-\lambda s}\, dL_{\lambda s} \stackrel{\mathrm{d}}{=} X.$$

Hence, we can write $X \stackrel{\mathrm{d}}{=} e^{-\lambda t} X_0 + W_{t\lambda}$, where $X_0 \stackrel{\mathrm{d}}{=} X$ with

$$W_{t\lambda} \stackrel{\mathrm{d}}{=} \int_{(0,t]} e^{-\lambda(t-s)}\, dL_{\lambda(t-s)}.$$

See Jurek and Mason [51] or Barndorff-Nielsen and Shephard [12]. The process L in (2.25) is termed the background Lévy driven process (BDLP) corresponding to the process X. An immediate consequence of Theorem 2.7 is that the limiting distribution for an OU-type process is self-decomposable.

In Chapter 5 we will focus on purely non-negative OU-type processes, which implies that the Gaussian component in the BDLP has to be zero.

An important and useful result involving the stochastic integral $\int h(u) dL_t$ is given by the following proposition:

Proposition 2.4. Let L = $\{L(t); t \geq 0\}$ be a Lévy process and $I := \int h(u)\, dL_t$. The characteristic functional for I can be computed by

$$\log \mathcal{C}_I(\xi) = \int \log \mathcal{C}_{L(1)}\left(\xi h(y)\right)\, dy. \tag{2.30}$$

For a proof of this proposition see Lukacs [68]. Using this result, Wolfe [107] established condition (2.26).

The main component when working with SD distributions or OU-type processes is the description of the innovation variable, namely W_ρ or W_t. This quantity becomes essential in issues like estimation, forecasting or simulation of SD distributions or OU-type processes. In Chapter 5 we explore a novel approach to analyze such innovation variables.

CHAPTER 3

LATENT STRUCTURE BASED MODELS: THEORY

This chapter introduces the general idea used in this thesis to construct stationary models with invariant distributions belonging to a specific parametric family. The first four sections present an introductory discussion, some relevant literature review and motivation. Section 3.5 formalizes the construction at issue. An alternative approach for this construction is described in Section 3.6. Section 3.7, is concerned with estimation schemes to be considered in this thesis. In the finial part, Section 3.8, a summary highlighting some relevant points is provided.

3.1 INTRODUCTION

In many situations where time-indexed data are studied, the question whether the underlying models have an invariant law possessed by the phenomenon is of interest. For instance, within the time-series analysis framework, strict stationarity conditions or less restrictive covariance stationary conditions (where only the first two moments are considered) are commonly a focus of research. Also, when working with Markov processes in continuous time, such as diffusion processes, the existence of finite invariant measures ensures some stability properties for the underlying diffusion process and at the same time enables some applicability, for example, to interest rate models in finance.

Under the uncertainty about the stationarity of the process in study, it is reasonable to work, from the outset, with a more general model that allows as a particular case

a stationary model. On the other hand, if we are able to tell from the outset that the model is stationary with invariant distribution belonging to a specific parametric family, that is, we accept that some characteristic properties of the underlying process do not change when time goes by, then in this case, the attention should be focused to the set of possible models with such stationary behavior.

There exists a vast literature concerned with conditions for stationarity and properties of stationary models. Depending on the area of interest, namely Markov chains, time series, diffusion processes etc. many results are available. See for instance, Cramér and Leadbetter [25] and Gikhman and Skorokhod [40]. On the other hand, the literature dedicated to the construction of suitable models with specific families of invariant distributions is more limited. As we mentioned before, this thesis is mainly devoted to analyze and give some generalizations of an idea introduced by Pitt et al. [78, 79] that allows the construction of models with specific stationary distributions. Before introducing this idea a literature review and motivation are given.

3.2 LITERATURE REVIEW

Much of the effort devoted towards the construction of stationary models with known invariant distributions, has been mainly focused on the discrete-time case. As far as this case is concerned, many papers dating from the late seventies have introduced stationary time series models, with invariant distributions belonging to specific parametric families. In particular, much of the attention has been given to autoregressive models.

To some extent the first specific accounts of the construction of stationary time series models, with non-Gaussian invariant distributions, were given by Lawrance and Lewis [60] and Jacobs and Lewis [44]. There, moving average models and autoregressive-moving average models with invariant distributions being exponentially distributed, were presented. The general exposition of these models resulted in an exponential autoregressive-moving average (EARMA(p, q)) model presented in Lawrance and Lewis [61]. In addition some other models with invariant distributions belonging to non-Gaussian specific parametric families were introduced, among these: Gaver and Lewis [39] and Lawrance [58] for mixed exponential distributions and for gamma distributions, McKenzie [72, 73] for negative binomial and Poisson distributions, Joe [48] and

Jørgensen and Song [50] for infinitely divisible convolution-closed distributions. Some other accounts can be found in Lawrance [59], Jørgensen and Song [50] and the references therein.

A relatively general approach to construct stationary autoregressive models, with invariant distribution in the family of convolution-closed infinitely divisible distributions, has been recently introduced by Joe [48] and Jørgensen and Song [50]. This approach plays an intuitive role for some of the constructions presented in this thesis. Here, we briefly review their approach.

Let η be a σ-finite positive measure on $\mathcal{E}$, where $\mathcal{E}$ denotes the Borel set of $\mathbb{R}$. Consider a random variable Z having probability density, with respect to η, given by

$$f_Z^\theta(z;\tau) = c(z;\tau)\,\exp\{z\theta - \tau k(\theta)\}, \tag{3.1}$$

where $k(\theta) = \log \int_{\mathbb{R}} c(z;\tau)\, e^{\theta z}\, d\eta/\tau$, $\theta \in \Theta$ with $\Theta = \text{int}\{\theta \in \mathbb{R}^d; k(\theta) < \infty\}$ and Θ is assumed to be non-empty. For $\tau = 1, 2, \ldots$, the probability density (3.1) characterizes the distribution of the sum $\sum_i^\tau X_i$, where $X_1, \ldots, X_\tau$ is a sample from the natural exponential family, see Barndorff-Nielsen [8]. Here, we assume that $\tau \in \mathcal{R}_+$ so that Z is infinitely divisible. Following Jørgensen and Song [50], we denote a random variable Z with density (3.1) as $Z \sim \text{ED}^*(\theta;\tau)$, where ED stands for ***exponential dispersion***. Notice that, if $Z \sim \text{ED}^*(\theta,\tau)$, then $\mathbb{E}[e^{\xi Z}] = e^{\tau\{k(\theta+\xi)-k(\theta)\}}$ and consequently

$$\mu = \mathbb{E}[Z] = \tau\, k'(\theta) \quad \text{and} \quad \text{Var}[Z] = \tau\, k''(\theta),$$

provided Z has second moment. Model (3.1) can more precisely be referred to as an infinitely divisible exponential dispersion model.

In general, if we have two independent random variables Y_1, Y_2 and define $X = Y_1 + Y_2$ then it follows that

$$F_X(x) = \int_{-\infty}^{\infty} F_{X|Y_1}(x \mid y_1)\, F_{Y_1}(dy_1) \tag{3.2}$$

$$= \int_{-\infty}^{\infty} F_{Y_2}(x - y_1)\, F_{Y_1}(dy_1). \tag{3.3}$$

Notice that $F_{X|Y_1}(x \mid y_1) = F_{Y_2}(x - y_1)$. Therefore, if we assume $Y_i \sim \text{ED}^*(\theta;\tau_i)$,

$i = 1, 2$, then in terms of the probability densities, we have the following

$$\begin{aligned} f_X^\theta(x;\tau) &= \int_{-\infty}^{\infty} f_{Y_2}^\theta(x - y_1;\tau_2)\, f_{Y_1}^\theta(y_1;\tau_1)\, d\eta(y_1) \qquad (3.4) \\ &= \exp\{x\theta - \tau k(\theta)\} \int_{-\infty}^{\infty} c(x - y_1;\tau_2)\, c(y_1;\tau_1)\, d\eta(y_1) \\ &= c(x;\tau)\, \exp\{x\theta - \tau k(\theta)\}, \qquad (3.5) \end{aligned}$$

where $\tau = \tau_1 + \tau_2$. Clearly, the family of exponential dispersion models is closed under convolutions, that is, $\mathrm{ED}^*(\theta;\tau_1) * \mathrm{ED}^*(\theta;\tau_2) = \mathrm{ED}^*(\theta;\tau_1 + \tau_2)$. From equations (3.4) and (3.5) we see that

$$\begin{aligned} f_{Y_1|X}(y_1 \mid x;\tau_1,\tau_2) &= \frac{f_{Y_2}^\theta(x - y_1;\tau_2)\, f_{Y_1}^\theta(y_1;\tau_1)}{f_X^\theta(x;\tau)} \\ &= \frac{c(x - y_1;\tau_2)\, c(y_1;\tau_1)}{c(x;\tau)}. \qquad (3.6) \end{aligned}$$

Here, we have dropped the dependence on θ since $f_{Y_1|X}$ no longer depends on θ. A way to specify an exponential dispersion model is by specifying the function $c(\cdot,\tau)$ or equivalently by specifying the function $k(\theta)$. A distribution with density (3.6) is denoted by $G(\tau_1,\tau_2,x)$ and is sometimes referred to as the contraction corresponding to $\mathrm{ED}^*(\theta;\tau)$. A key property to point out when using ED models in the convolution scheme described above is that Y and $X - Y$ are independent with $Y \sim \mathrm{ED}^*(\theta;\tau_1)$ and $X - Y \sim \mathrm{ED}^*(\theta;\tau_2)$.

Joe [48] considered the more general class of convolution-closed infinite divisible distributions which includes distributions without moments. However, as pointed out in [50], these distributions can be seen as the family of exponential dispersion models with $\theta = \{0\}$, that is $\{\mathrm{ED}^*(0;\tau);\tau > 0\}$. Here, we state the construction of Joe [48] when the underlying models are defined through (3.1). See Jørgensen and Song [50].

Assume $\tau_1 = \rho\lambda$, $\tau_2 = (1-\rho)\lambda$ and $0 < \rho < 1$ then $\tau = \lambda$. The stationary AR(1) model introduced in Joe [48] is given by

$$X_t = A_t(X_{t-1},\rho) + \varepsilon_t, \quad t = 1, 2, \ldots \qquad (3.7)$$

Where $A_t(\cdot,\rho)$ is a random operator, independent from the innovation variable ε_t,

defined as follows:

$$A_t(X,\rho) \mid [X = x] \sim G(\rho\lambda, (1-\rho)\lambda, x) \tag{3.8}$$

and marginally $A_t(X,\rho) \sim \text{ED}^*(\theta;\rho\lambda)$. The innovations ε_t are IID with distribution $\text{ED}^*(\theta;(1-\rho)\lambda)$. The marginal distribution for the AR(1) model (3.7) has distribution $\text{ED}^*(\theta;\lambda)$.

Example 3.1. Let us specify the model through $k(\theta) = \theta^2/2$, then

$$\int_{\mathbb{R}} c(x,\lambda)\, e^{x\theta}\, dx = e^{\lambda\frac{\theta^2}{2}},$$

which in turn implies that $c(x,\lambda) = \text{N}(x;0,\lambda)$. With this quantity the exponential dispersion model is completely characterized. Hence, the innovations in the AR(1) model (3.7) are given by

$$\varepsilon_t \sim \text{ED}^*(\theta;(1-\rho)\lambda) = \text{N}(\theta(1-\rho)\lambda, (1-\rho)\lambda).$$

Following equation (3.6), we get $G(\rho\lambda,(1-\rho)\lambda,x) = \text{N}(\rho x, \rho(1-\rho)\lambda)$. In a similar way we can obtain $\text{ED}^*(\theta;\rho\lambda) = \text{N}(\theta\rho\lambda,\rho\lambda)$ and therefore we can rewrite the convolution formula (3.4) as

$$\text{N}(\theta\rho\lambda,\rho\lambda) * \text{N}(\theta(1-\rho)\lambda,(1-\rho)\lambda) = \text{N}(\theta\lambda,\lambda),$$

from where is clear that the marginal or stationary distribution for the AR(1) model (3.7) is given by $\text{N}(\theta\lambda,\lambda)$. The dynamics for this model can by written through the following equation

$$X_t = \rho X_{t-1} + \omega_t,$$

where ω_t are IID with distribution $\text{N}(\theta(1-\rho)\lambda,(1-\rho^2)\lambda)$. ∘

In the same way as in Example 3.1, many stationary time series models with invariant distribution being ED were introduced in [48] and [50].

A different point of view of convolution closed exponential dispersion models is given by the relation to Lévy processes. That is, if we assume that $X = \{X_t; t \geq 0\}$ is a

Lévy process, then the density (3.1) characterizes the distribution of the underlying increments $X_{\tau+u} - X_u$. The fascinating theory behind the latter connection is treated in Küchler and Sørensen [55, 56].

Whereas the approaches mentioned so far provide with techniques to construct stationary models with specific invariant distributions, there is not, to the best of our knowledge, a general way to estimate the parameters of such models.

So far, all the models we have mentioned belong to the class of discrete-time models, more precisely, to the class of time series models. For the continuous-time case (more general), the literature dedicated to the construction of stationary models with given invariant distribution is even more limited. From a constructive point of view, we may say that, the construction of stationary processes has been mainly reduced to covariance stationary processes. That is, by specifying some homogeneity conditions to the mean and covariance functions. This in turn leads to the study of (strictly) stationary Gaussian process, since in this case they are completely characterized by the mean and covariance functions. The theory available on stationary Gaussian processes is relatively wide and is available in many text books, see for example Cramér and Leadbetter [25] and Ibragimov and Rozanov [43]. See also Example 2.2. As in the discrete-time case much of the theory has been focused on the Markov case, which in the Gaussian case leads to Ornstein-Uhlenbeck processes. The difficulty here arises when the interest is to construct continuous-time stationary models with invariant distributions outside the Gaussian family. In this line, an important contribution can be found in Barndorff-Nielsen et al. [11, 13] and the references therein. There, the construction of stationary processes in continuous time, with invariant distributions being self decomposable, is available via OU-type processes. A review of these processes is given in Section 2.6. Chapter 5 provides with a description of the innovation part corresponding to SD distributions, that eases the study of OU-type processes.

Another approach to construct continuous-time stationary models has been done for one-dimensional diffusion processes. Here the methodology used to construct the underlying drift and diffusion coefficients such that the resulting diffusion is stationary, has been mainly based on a relation of the invariant measure with the scale and the speed measure similar to that reviewed at the end of Section 2.3. Within this approach, Yacine [108] estimated the invariant density of a stationary diffusion and Rydberg [88]

introduced a diffusion model with stationary distribution being generalized hyperbolic. A drawback of the above models is that the transition density corresponding to the constructed diffusion is not always known, leading to more complicated estimation procedures. More recently, Bibby and Sørensen [15] have described a constructive way of defining diffusion models with invariant distributions on the positive real line. The latter approach has the characteristic property that the underlying autocorrelation structure is known. However, estimation procedures based on the full density such as maximum likelihood estimation are not directly available. As we will see the approach described in this chapter will allow us to know, when available, the whole transition probability.

3.3 TIME REVERSIBILITY

An important property to underline here, satisfied for most of the models mentioned in Section 3.2, is the property of time reversibility. For example, time reversibility is satisfied by the autoregressive model (3.7), Joe [48], and all stationary Gaussian processes, Weiss [104].

Definition 3.1. A stochastic processes $\mathrm{X} = \{X(t); t \in \mathcal{T}\}$ is said to be ***time reversible*** if

$$\{X_{t_1}, X_{t_2}, \ldots, X_{t_n}\} \stackrel{\mathrm{d}}{=} \{X_{t_n}, X_{t_{n-1}}, \ldots, X_{t_1}\} \tag{3.9}$$

holds for any $t_1, \ldots, t_n \in \mathcal{T}$ and $n = 1, 2, \ldots$

Notice that time reversibility is a distributional property rather than a sample path property. The sample path counterpart is often referred to as the ***time reversal*** process. That is, we can say that a process is time reversible if the time reversed process has the same law as the original one. See Rogers and Williams [84] for more on time reversal.

Time reversibility may be seen as an unsuitable property to model data that show certain clear asymmetry with respect to a given time point, such as the classical yearly sunspot data. See Lawrance [59] for a review of reversibility in time series. However, from a modelling point of view, time reversible models play an important role, for

example it can be shown that practically all one dimensional diffusions are reversible, see Kent [54].

Definition 3.1 is a general form of reversibility, however, other restricted definitions may arise. For example, ***lag reversibility*** where the condition

$$\{X_{t+r}, X_t\} \stackrel{\mathrm{d}}{=} \{X_t, X_{t+r}\}, \quad r = 1, 2, \ldots$$

is required. Lawrance [59] noticed that lag-reversibility implies marginal stationarity and that

$$\{X_{t+r} - X_t\} \stackrel{\mathrm{d}}{=} \{X_t - X_{t+r}\}.$$

This, also leads to the conclusion that $\{X_{t+r} - X_t\}$ has a symmetric distribution and, consequently, $P[X_t < X_{t+r}] = P[X_t > X_{t+r}] = 1/2$, so the probability of run-up behavior is the same as the probability of run-down behavior. Lawrance [59] also noticed that another consequence of lag-reversibility is

$$\mathrm{Corr}(X_t, X_{t+r}^2) = \mathrm{Corr}(X_t^2, X_{t+r}) = \mathrm{Corr}(X_t, X_{t-r}^2).$$

Clearly, time reversibility is a higher order property of dependence.

We pose ourselves the task of constructing stationary models by first imposing the time reversibility as a characteristic rather than as a property. We simplify this task by mainly focusing on time homogeneous Markov processes in discrete and continuous time.

The idea of using time reversibility as characteristic property was first introduced by Pitt et. al. [78] and Pitt and Walker [79]. There, only the discrete-time case was treated. In what follows we give a general motivation, inspired on some of the ideas presented in [78] and [79]. Further accounts of these references will be reviewed in Chapter 4.

3.4 MOTIVATION AND GENERAL PROBLEM

Suppose that we are in the situation where, from the outset, we know that the time-indexed phenomenon under study has a known stationary distribution and consequently

we are interested in models with such a stationary distribution. For the sake of illustration, let us assume that we are interested in discrete time-homogeneous models with the Markov property being satisfied, we will treat the continuous case (more general) later. We denote a generic sequence from such a model by $\mathrm{X} = \{X_t\}_{t=1}^{\infty}$.

It is well known that the *fdds* corresponding to a Markov process are determined by its corresponding transition probabilities and initial measure, see equation (2.9). Assuming that absolutely continuous densities for the transition probability and the initial probability exist, we can write the *fdds* of a discrete time-homogeneous Markov process as follows,

$$p(x^{(t,n)}) = q(x_t) \prod_{i=1}^{n} p(x_{t+i} \mid x_{t+i-1}) \tag{3.10}$$

where $x^{(t,n)} = (x_t, \ldots, x_{t+n})$, $t, n \in \mathbb{N}$ and q and p denote the densities for the initial measure Q and the one-step transition probability P respectively. Due to the time-homogeneity of the transition probability, the analysis of (3.10) is reduced to the analysis of the joint distribution $f_{X_{t+1},X_t}(x_{t+1}, x_t)$, or $f_{X_{t+r},X_t}(x_{t+r}, x_t)$, $r = 1, 2, \ldots$ when using r-step transition probabilities to construct the *fdds*.

One way to ensure that the process X preserves the distribution $Q(\cdot)$ as its marginal, can be given by assuming certain symmetry on the underlying joint distribution. This symmetry can be achieved by assuming

$$f_{X_{t+1},X_t}(x_{t+1}, x_t) = f_{X_{t+1},X_t}(x_t, x_{t+1}), \tag{3.11}$$

for all $t = 1, 2, \ldots$ That is, the two-dimensional distributions are invariant under permutations. Equality (3.11) is sometimes referred to as ***first-order time reversibility***. This is mainly the kind of reversibility treated in this thesis, which implies the general definition (3.9) in the case of time-homogeneous Markov processes. In fact (3.11) can be ensured by assessing the well-known detailed balance condition. For instance, we could assume

$$f_{X_{t+1},X_t}(x_{t+1}, x_t) = p(x_{t+1} \mid x_t)\, q(x_t) = p(x_t \mid x_{t+1})\, q(x_{t+1}). \tag{3.12}$$

For certain transition density p and given initial density q.

Here, we do not want to drop the dependence on X_t, hence we exclude IID sequences as a possible construction of the required *fdds*. The time reversibility condition (3.11) gives a way to construct the *fdds* without dropping the lag-dependence on X_t in the transition $\mathrm{P}(\cdot \mid X_t = x_t)$. Clearly a suitable parametric form for this transition must be found. Hence, a method to generate reversible sequences must lead to transition probabilities $\mathrm{P}(\cdot \mid X_t = x_t)$ that stay invariant under $Q(\cdot)$.

With a different aim, the statistics literature provides many ways of constructing reversible sequences via MCMC methods such as the Gibbs sampler and Metropolis-Hasting algorithms. See Robert and Casella [82] and the references therein for more on these methods. Let us take as an example the two-component Gibbs sampler method, which is based on the existence of an augmented joint distribution $F_{X,Y}$ with conditionals $F_{X|Y}$ and $F_{Y|X}$. The one-step transition probability $\mathrm{P}(X_{t+1} \mid X_t)$ can be constructed via a Gibbs sampler type updating mechanism by simulating a latent process $\{Y_t\}_{t=1}^{\infty}$, that is

$$Y_{t+1} \mid X_t \sim F_{Y|X}(\cdot \mid X_t)$$

and

$$X_{t+1} \mid Y_{t+1} \sim F_{X|Y}(\cdot \mid Y_{t+1}).$$

The two-component Gibbs sampler has the property that the constructed sequences $\{X_t\}_{t=1}^{\infty}$ and $\{Y_t\}_{t=1}^{\infty}$ are reversible, see Liu et al. [65]. Discrete-time transition densities constructed in this way were introduced in Pitt et al. [78, 79] as a method to construct stationary first order autoregressive models and first order autoregressive conditional heteroscedastic models.

3.5 GIBBS SAMPLER TYPE CONSTRUCTION

The goal of this section is to lay the foundations of the construction introduced in Section 3.4 as well as to give its general properties. Most of the models studied in this thesis will have a time-homogeneous Markovian structure, therefore we set the Markov property as the priority in the general discussion given below.

Consider the following ***general setting***:

Given a measure $Q_X(\mathrm{E}) > 0$ on the Borel space $(\mathrm{E}, \mathcal{E})$ we want to construct a time-

homogeneous Markov process $\mathrm{X} = \{X(t); t \in \mathcal{T}\}$ (henceforth referred to as ***target process***) on $(\Omega, \mathcal{A}, \mathrm{P})$ with state space $(\mathrm{E}, \mathcal{E})$ and having the given measure Q_X as its invariant measure. In this thesis $\mathcal{T} = \mathbb{R}_+$ or $\mathcal{T} = \mathbb{N}_0$, when treating both cases generically we keep the notation $\mathcal{T}$. With these conditions we are able to work with the separable modification of X. Hence, when we speak about the target process X, the separable modification must be understood. See Section 2.1 for a definition of separable processes. When $\mathcal{T} = \mathbb{R}_+$, that is in the continuous time case, further continuity conditions must added to the target process, in general we assume that the process is standard.

Since we work with the separable modification then we can specify the process X through its *fdds* which in turn, for time-homogeneous Markov processes, lead us to the specification of the transition probabilities and the initial measure. Hence, given Q_X we only need to specify a valid transition probability $\mathrm{P}_t(x, B)$, $x \in \mathrm{E}$, $B \in \mathcal{E}$ which stays invariant under Q_X. That is,

$$Q_X(B) = \int_E \mathrm{P}_t(x, B)\, Q_X(dx). \tag{3.13}$$

Equality (3.13) is ensured to be reversible with respect to Q_X (or sometimes referred to as self-adjoint)[1]. In terms of the transition probabilities of a time-homogeneous process this translates to ensure the following condition

$$\int_B \mathrm{P}_t(x, B')\, Q_X(dx) = \int_{B'} \mathrm{P}_t(x, B)\, Q_X(dx), \tag{3.14}$$

for all sets $B, B' \in \mathcal{E}$. Notice that in some texts the reversibility condition (3.14) requires that $Q_X(\mathrm{E}) < \infty$, that is Q_X is a finite measure. However we do not insist on this and refer to the general case as Q_X-symmetric transition probabilities, see Silverstein [95]. If (3.14) is satisfied, then we say that P_t is a ***Q_X-reversible transition probability***.

As we have mentioned before, a way to construct $\mathrm{P}_t(x, B)$ such that (3.14) is satisfied can be done via a Gibbs sampler type update. We proceed as follows, first let us assume that $Q_X \ll \lambda_X$ and $\mathrm{P}_t \ll \lambda_X$ where λ_X is a σ-finite measure on $(\mathrm{E}, \mathcal{E})$, we will refer to their respective densities as q_X and p_t. As the process X is required to have Q_X

[1] In the statistics literature, the reversibility condition is sometimes called balance condition.

as an invariant measure, then we can say that marginally $X_t \sim Q_X$ for each $t \in \mathcal{T}$, provided Q_X is a probability measure[2]. Then, we can use the notation X, $X \stackrel{\mathrm{d}}{=} X_t$, when speaking marginally of X_t.

A completion $F_{X,Y}$ of Q_X can be constructed by introducing a $(\mathrm{S}, \mathcal{S})$-measurable random variable Y (henceforth referred to as the ***latent structure***) as follows:

$$Q_X(B) = \int_B \int_S F_{X,Y}(dx, dy),$$

where $F_{X,Y}$ is a measure on the product space $(\mathrm{E}\times\mathrm{S}, \mathcal{E}\otimes\mathcal{S})$. Notice that this completion is not unique. If $F_{X,Y}$ is a finite measure, then we can normalize it and we speak of a joint probability measure. If λ_Y denotes a σ-finite measure on $(\mathrm{S}, \mathcal{S})$ and $F \ll \lambda$, where $\lambda = \lambda_X \times \lambda_Y$, then we denote the corresponding density as f. Let us briefly suppose that Q_X is finite such that a probability measure can be associated via normalization. Since we know the marginal for X, in order to construct a completion $F_{X,Y}$, it is enough to provide a valid conditional probability $F_{Y|X}$, which in terms of the densities reduces to

$$f_{X,Y}(x, y) = f_{Y|X}(y \mid x)\, q_X(x). \tag{3.15}$$

In order to construct a time reversible transition $\mathrm{P}_t(x, B)$, we introduce a latent process $\{Y(t); t \in \mathcal{T}\}$ on $(\Omega, \mathcal{A}, \mathrm{P})$ with Borel state-space $(\mathrm{S}, \mathcal{S})$ through the following Gibbs sampler type updating mechanism

$$\begin{aligned} \{Y_t \mid X_0 = x_0\} &\sim F_{Y|X}(\cdot \mid x_0) \\ \{X_t \mid Y_t = y_t\} &\sim F_{X|Y}(\cdot \mid y_t), \end{aligned} \tag{3.16}$$

where y_t is used to denote a space-time point in $\mathrm{S} \times \mathcal{T}$. The conditional distribution $F_{X|Y}$, also required in (3.16), can be found by an application of Bayes theorem. In the above conditional distributions we have omitted the notation corresponding to the dependence on t for notation simplicity. However, we will frequently use the subindex t in the conditional distributions, for example $F_{Y_t|X_0}$ or $F_{Y_{t+1}|X_t}$, to emphasize the stage of the updating in which we are situated. Notice that introducing a latent process, as in

[2]Even when Q_X is not a probability measure, we abuse the term "marginal" to refer to the measure associated to X_t for all $t \in \mathcal{T}$.

(3.16), imposes the following conditional independence structure between $\{X(t); t \in \mathcal{T}\}$ and $\{Y(t); t \in \mathcal{T}\}$

$$\begin{aligned} \mathrm{P}\left[Y_t \in C \mid \mathcal{X}_s, \mathcal{Y}_s\right] &= \mathrm{P}\left[Y_t \in C \mid \sigma(X_s)\right] \\ \mathrm{P}\left[X_t \in B \mid \mathcal{Y}_t, \mathcal{X}_s\right] &= \mathrm{P}\left[X_t \in B \mid \sigma(Y_t)\right], \quad s \leq t; s, t \in \mathcal{T}, \end{aligned} \tag{3.17}$$

for all $B \in \mathcal{E}$, $C \in \mathcal{S}$ and where $\mathcal{Y}_s = \sigma(Y_u; u \leq s)$, $\mathcal{X}_s = \sigma(X_u; u \leq s)$, $\mathcal{Y}_s, \mathcal{X}_s \subseteq \mathcal{A}$. In the discrete-time setting, equalities (3.17) can be interpreted as the interleaving property for the two component Gibbs sampler defined in Liu et al. [65].

Using (3.17) we see that,

$$\mathrm{P}(X_t \in B \mid \mathcal{X}_s) = \int_S P(X_t \in B \mid Y_t = y)\, P(dy \mid \mathcal{X}_s) = \mathrm{P}(X_t \in B \mid \sigma(X_s)). \tag{3.18}$$

Therefore, for a given value $X_0 = x$, we can define the transition probability given by

$$\mathrm{P}_t(x, B) := \int_S F_{X_t|Y_t=y}(B)\, F_{Y_t|X_0=x}(dy), \quad x \in \mathrm{E}, B \in \mathcal{E}, \tag{3.19}$$

where $\mathrm{P}_t(x, B) = \mathrm{P}\left[X_t \in B \mid X_0 = x\right]$ and

$$F_{X_t|Y_t=y}(B) := \mathrm{P}[X_t \in B \mid Y_t = y], \quad B \in \mathcal{E}, \tag{3.20}$$

$$F_{Y_t|X_0=x}(C) := \mathrm{P}[Y_t \in C \mid X_0 = x], \quad C \in \mathcal{S}. \tag{3.21}$$

The transition probability $\mathrm{P}_t(x, B)$ defines a time-homogeneous Markov transition probability with state space $(\mathrm{E}, \mathcal{E})$ provided it satisfies Chapman-Kolmogorov equations (2.5). With this transition probability and an initial (prior) measure, Kolmogorov's existence theorem ensure the existence of a version of a Markov process $\{X(t); t \in \mathcal{T}\}$.

An important point to underline here is that, in order to have a well defined Markov model, the transition probability (3.19) must satisfy Chapman Kolmogorov equations. In the discrete-time case, as in the case of the well known Gibbs sampler method, these equations are immediately satisfied, see Robert and Cassella [82]. However, in the continuous time case, Chapman Kolmogorov equations are not so straightforward since the time information may come in many ways, see Example 3.3 below. Therefore, suitable ways to incorporate the time information must be found such that these equations are

satisfied. This, is basically the splitting point between the discrete-time case treated in Chapter 4 and the continuous-time case treated in Chapter 6.

Also notice that in contrast to the usual Gibbs sampler method, which is commonly not started from the stationary distribution, the generated process in the construction at issue, is always in stationarity due to the fact that we have given the invariant distribution.

Proposition 3.1. The transition probability constructed via the following density

$$p_t(x_t \mid x_0) = \int_S f_{X|Y}(x_t \mid y_t)\, f_{Y|X}(y_t \mid x_0)\, \lambda_Y(dy_t) \tag{3.22}$$

is reversible for any choice of q_X and $t \in \mathcal{T}$.

Proof.

$$\begin{aligned} q_X(x_0)\, p_t(x_t \mid x_0) &= q_X(x_0) \int_S f_{X|Y}(x_t \mid y_t)\, f_{Y|X}(y_t \mid x_0)\, \lambda_Y(dy_t) \\ &= \int_S f_{X|Y}(x_t \mid y_t)\, f_{Y,X}(y_t, x_0)\, \lambda_Y(dy_t) \\ &= q_X(x_t) \int_S f_{Y|X}(y_t \mid x_t)\, f_{X|Y}(x_0 \mid y_t)\, \lambda_Y(dy_t) \\ &= q_X(x_t)\, p_t(x_0 \mid x_t). \end{aligned} \tag{3.23}$$

□

When $\mathcal{T}$ is countable, Proposition 3.1 is a well known property of a Markov chain constructed using the Gibbs sampler method. See Liu et al. [65].

In order to illustrate the above construction we give two examples the first constitutes an autoregressive model, that is a model in discrete time, whereas the second represent a continuous-time Markov chain.

Example 3.2. Suppose we are studying a model evolving in time $t = 1, 2, \ldots$ for which we know or assume that the stationary distribution is given by $\mathrm{N}(0, 1)$. We can construct a Markov process $\{X_t\}_{t=1}^{\infty}$ with the required stationary distribution as follows:

Let us introduce the random variable Y via the joint density given by

$$f_{Y,X}(y, x) = f_{Y|X}(y \mid x)\, q_X(x),$$

where $X \sim \mathrm{N}(0,1)$. With this joint density it is possible to construct the following Gibbs type update

- $\{Y_{t+1} \mid X_t = x\} \sim F_{Y|X}(\cdot \mid x)$
- $\{X_{t+1} \mid Y_{t+1} = y\} \sim F_{X|Y}(\cdot \mid y)$,

where $\{Y_t\}_{t=1}^{\infty}$ represents another process termed the latent or hidden process. In this case, the independence structure (3.17) rewrites as

$$\mathrm{P}\left(Y_{t+1} \mid \mathcal{Y}^{(t)}, \mathcal{X}^{(t)}\right) = \mathrm{P}(Y_{t+1} \mid \sigma(X_t)) \tag{3.24}$$

$$\mathrm{P}\left(X_{t+1} \mid \mathcal{Y}^{(t+1)}, \mathcal{X}^{(t)}\right) = \mathrm{P}(X_{t+1} \mid \sigma(Y_{t+1})) \tag{3.25}$$

where $\mathcal{X}^{(t)} := \sigma(X_1, X_2, \ldots, X_t)$. [3]

If we assume $F_{Y|X}(\cdot \mid x) = \mathrm{N}(x, 1)$, then $F_{X|Y}(\cdot \mid y) = \mathrm{N}(y/2, 1/2)$ and following (3.19), the resulting one-step transition probability is given by

$$\mathrm{P}(x_t, B) = \int_B p(x_t, x_{t+1})\, dx_{t+1},$$

where

$$\begin{aligned} p(x_t, x_{t+1}) &= \int_{-\infty}^{\infty} \mathrm{N}(x_{t+1}; y/2, 1/2)\, \mathrm{N}(y; x_t, 1)\, dy \\ &= \int_{-\infty}^{\infty} \frac{\exp\left\{-(x_{t+1} - \frac{y}{2})^2 - \frac{1}{2}(y - x_t)^2\right\}}{\sqrt{\pi}\sqrt{2\pi}}\, dy \\ &= \int_{-\infty}^{\infty} \frac{\sqrt{6}\exp\left\{-\frac{2}{3}(x_{t+1} - x_t/2)^2\right\}}{3\sqrt{\pi}} \\ &= \mathrm{N}(x_{t+1}; x_t/2, 3/4). \end{aligned}$$

Using this transition probability is possible to construct a version of a Markov process with invariant distribution $\mathrm{N}(0,1)$.

In the same way, if instead of $\mathrm{N}(x,1)$, we assume $\mathrm{N}(x,s)$, $s > 0$, as the distribution corresponding to $F_{Y|X}(\cdot \mid x)$. In this case, the resulting transition probability is found to be

$$\mathrm{P}(X_{t+1} \mid X_t = x_t) = \mathrm{N}\left(\frac{x_t}{s+1}, \frac{(2+s)s}{(s+1)^2}\right).$$

[3] When regarded as model assumptions, equalities (3.24) and (3.25) are known as the state distribution and the observation distribution of an observation-driven model. See Cox [22].

The corresponding Markov process to the above transition has also N(0, 1) as an invariant probability. Therefore, for each $s > 0$ it is possible to construct a Markov process with the same invariant measure. ∘

The above example can be seen as a particular case of Example 3.1 if we allow $\theta = 0$ and fix $\lambda = 1$ and $s = \rho^{-1} - 1$. More generally, Pitt et al. [78] proved that the approach described by Jørgensen and Song [50] can be seen as a particular case of the construction at issue.

Example 3.3. Suppose we want to construct a continuous-time Markov process $\mathrm{X} = \{X_t; t \in \mathbb{R}_+\}$ with invariant distribution Po(1), that is, $q_X(x) = e^{-1}/x!$, $x = 0, 1, \ldots$ Also let us assume that

$$f_{Y|X}(y \mid x; \xi) = \mathrm{Bi}(y; x, 1-\xi) = \binom{x}{y} \xi^{x-y} \, (1-\xi)^y \, \mathbb{I}_{\{0,\ldots,x\}}(y), \quad 0 < \xi < 1.$$

After an application of Bayes theorem we get

$$f_{X|Y}(x \mid y; \xi) = \frac{\xi^{x-y}}{(x-y)!} \, e^{-\xi} \mathbb{I}_{\{y,\ldots,\infty\}}(x).$$

Therefore, we construct the required process X by introducing a latent structure $\mathrm{Y} = \{Y_t; t \in \mathbb{R}_+\}$ via the updating mechanism

$$\begin{aligned} \{Y_t \mid X_0 = x_0\} &\sim f_{Y|X}(\cdot \mid x_0; \xi_t) \\ \{X_t \mid Y_t = y_t\} &\sim f_{X|Y}(\cdot \mid y_t; \xi_t), \end{aligned}$$

where $\xi_t = 1 - e^{-t}$.

In this case the transition probability (3.19) has mass function given by

$$\begin{aligned} p_t(x_0, x_t) &= \sum_{y=0}^{\infty} f_{X|Y}(x_t \mid y; \xi_t) \, f_{Y|X}(y \mid x_0; \xi_t) && (3.26) \\ &= e^{-\xi_t} \, \xi_t^{x_0+x_t} \, x_0! \sum_{y=0}^{x_0 \wedge x_t} \frac{(\frac{1-\xi_t}{\xi_t^2})^y}{y! \, (x_0-y)! \, (x_t-y)!} \\ &= \frac{e^{-\xi_t} \, \xi_t^{x_0+x_t}}{x_t!} \sum_{y=0}^{x_0 \wedge x_t} \frac{(\frac{1-\xi_t}{\xi_t^2})^y}{y!} \, (-x_0)_y \, (-x_t)_y && (3.27) \\ &= \frac{e^{-\xi_t} \, \xi_t^{x_0+x_t}}{x_t!} \, {}_2F_0\left(-x_0, -x_t; \frac{1-\xi_t}{\xi_t^2}\right), && (3.28) \end{aligned}$$

where $a \wedge b$ stands for $\min\{a, b\}$ and ${}_2F_0$ is a generalized hypergeometric function, see Abramowitz and Stegun [2], formulas 15.4.1 and 15.4.2. In expression (3.27), $(a)_n$ denotes the Pochhammer function defined as $(a)_n = \Gamma(a+n)/\Gamma(a)$. Also notice that, for this expression, we have used the relation $1/(x-y)! = (-1)^y(-x)_y/x!$.

To ensure that the process X induced through the transition probability (3.28) is a well-defined Markov process, we need to verify Chapman-Kolmogorov equations (2.5). Let us denote $\mathcal{L}_X(\theta) = \mathbb{E}[e^{X\theta}]$ the Laplace transform corresponding to the random variable X. Hence, we find that

$$\begin{aligned} \mathcal{L}_{Y|X}(\theta) &= \left\{\xi_t + e^{\theta}(1-\xi_t)\right\}^x \quad \text{and} \\ \mathcal{L}_{X|Y}(\theta) &= e^{y\theta} e^{\xi_t(e^{\theta}-1)}. \end{aligned}$$

The Laplace transform corresponding to the transition mass function (3.28) can be computed as

$$\begin{aligned} \mathcal{L}_{X_t|X_0=x}(\theta) &= \mathbb{E}[\mathcal{L}_{X_t|Y_t}(\theta) \mid X_0 = x] \\ &= e^{\xi_t(e^{\theta}-1)} \mathcal{L}_{Y_t|X_0=x}(\theta) \\ &= e^{\xi_t(e^{\theta}-1)} \left\{\xi_t + e^{\theta}(1-\xi_t)\right\}^x. \end{aligned} \tag{3.29}$$

Using the Laplace transform (3.29), to satisfy the Chapman-Kolmogorov equations is equivalent to satisfy

$$\mathbb{E}\left[\mathcal{L}_{X_{t+s}|X_s}(\theta) \mid X_0 = x\right] = \mathcal{L}_{X_{t+s}|X_0=x}(\theta). \tag{3.30}$$

Hence, we verify equality (3.30) as follows

$$\begin{aligned} &\mathbb{E}[\mathcal{L}_{X_{t+s}|X_s}(\theta) \mid X_0 = x] \\ &= e^{\xi_t(e^{\theta}-1)} \mathcal{L}_{X_s|X_0=x}(\widehat{\theta}), \quad \text{where} \quad \widehat{\theta} = \log\left(\xi_t + e^{\theta}(1-\xi_t)\right) \\ &= e^{\xi_t(e^{\theta}-1)} \left[\exp\{\xi_s[\xi_t + e^{\theta}(1-\xi_t) - 1]\} \left\{\xi_s + [\xi_t + e^{\theta}(1-\xi_t)](1-\xi_s)\right\}^x\right] \\ &= \exp\{(\xi_t + \xi_s - \xi_t\xi_s)(e^{\theta}-1)\} \left[\xi_t + \xi_s - \xi_t\xi_s + e^{\theta}(1-\xi_t)(1-\xi_s)\right]^x \\ &= \exp\left\{\xi_{t+s}(e^{\theta}-1)\right\} \left[\xi_{t+s} + e^{\theta}(1-\xi_{t+s})\right]^x \qquad (3.31) \\ &= \mathcal{L}_{X_{t+s}|X_0=x}(\theta). \qquad (3.32) \end{aligned}$$

Notice that in equation (3.31) we have use the fact that $\xi_{t+s} = \xi_t + \xi_s - \xi_t\xi_s$ and $(1-\xi_{t+s}) = (1-\xi_t)(1-\xi_s)$, equalities that are satisfied for $\xi_t = 1 - e^{-t}$. However, is worth noting that the Chapman-Kolmogorov equations are not satisfied, for arbitrary choice of ξ_t.

Also using the the Laplace transform (3.29) it is straightforward to compute the following moments

$$\begin{aligned} \mathbb{E}[X_t \mid X_0 = x] &= \mathcal{L}'_{X_t|X_0}(0) = \xi_t + (1-\xi_t)\,x \quad \text{and} \\ \mathrm{Var}[X_t \mid X_0 = x] &= \mathcal{L}''_{X_t|X_0}(0) - \{\mathcal{L}'_{X_t|X_0}(0)\}^2 = \xi_t + (\xi_t - \xi_t^2)x. \end{aligned}$$

Notice that in this continuous-time setting the model has transition probabilities with linear expectation.

Within the continuous Markov chains framework, a process X with transition probabilities (3.28) can be recognized as a birth and death process. First let us notice that, in order to have standard transition probabilities it remains to check that $\lim_{t\downarrow 0} p_t(x,\cdot) = \delta_x(\cdot)$. This follows easily using the representation (3.26)

$$\begin{aligned} \lim_{t\downarrow 0} p_t(x_0, x_t) &= \sum_{y=0}^{\infty} \lim_{t\downarrow 0} f_{X|Y}(x_t \mid y; \xi_t)\, f_{Y|X}(y \mid x_0; \xi_t) \\ &= \sum_{y=0}^{\infty} \left\{ \lim_{t\downarrow 0} f_{X|Y}(x_t \mid y; \xi_t) \right\} \left\{ \lim_{t\downarrow 0} f_{Y|X}(y \mid x_0; \xi_t) \right\} \qquad (3.33) \\ &= \mathbb{I}(x_t = x_0), \end{aligned}$$

for any states $x_t, x_0 = 0, 1, \ldots$ The equality (3.33) follows due to the polynomial form of the conditional densities. With this, the infinitesimal generator $Q = \{q_{ij}\}$ of the semigroup $\mathrm{P}_t = \{p_{ij}(t)\}$, where $p_{ij}(t) := p_t(i,j)$, is given by

$$\begin{aligned} q_{ij} &= \begin{cases} -\lim_{t\downarrow 0} \frac{1-p_{ii}(t)}{t}, & i = j \\ \lim_{t\downarrow 0} \frac{p_{ij}(t)}{t}, & i \neq j \end{cases} \\ &= \begin{cases} -(i+1), & i = j \\ 1, & j = i+1 \\ i, & j = i-1 \\ 0, & \text{otherwise.} \end{cases} \end{aligned}$$

It can be seen that in this case the above process is strong Markov, since all the states are stable, that is $0 \leq -q_{ij} < \infty$.

In matrix notation the above infinitesimal generator can be written as

$$Q = \begin{bmatrix} -1 & 1 & 0 & 0 & 0 & \cdots \\ 1 & -2 & 1 & 0 & 0 & \cdots \\ 0 & 2 & -3 & 1 & 0 & \cdots \\ 0 & 0 & 3 & -4 & 1 & \cdots \\ 0 & 0 & 0 & 4 & -5 & \cdots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \ddots \end{bmatrix} \tag{3.34}$$

This infinitesimal generator is immediately recognized as the corresponding to a conservative birth and death process with birth rate $\lambda_i = 1$ and death rate $\mu_i = i$. All the classical results for this process follow, for example the waiting times in state i are exponential with rate $\lambda_i + \mu_i = 1 + i$. Also a well known result establish that the corresponding stationary distribution is given by

$$\begin{aligned} q_X(n) &= \frac{\lambda_0 \lambda_1 \cdots \lambda_{n-1}}{\mu_1 \mu_2 \cdots \mu_n} q_X(0), \quad n \geq 1 \\ &= \frac{1}{n!} q_X(0), \end{aligned}$$

where

$$q_x(0) = \left(\sum_{n=0}^{\infty} \frac{\lambda_0 \lambda_1 \cdots \lambda_{n-1}}{\mu_1 \mu_2 \cdots \mu_n} \right)^{-1} = \left(\sum_{n=0}^{\infty} \frac{1}{n!} \right)^{-1} = e^{-1},$$

as required. ∘

Notice that, in general, given an initial distribution and an infinitesimal generator Q is not possible to determine the complete behavior of a continuous-time Markov chain. However, in the example given above, $\sum_{i=0}^{\infty} -q_{ii}^{-1} = \infty$, so the process does not explode and therefore the association of our constructed process and the birth and death process is safe. Further considerations for the process after the explosion time must be set if the chain explodes.

Example 3.3 can be generalized to have a Poisson invariant measure with intensity parameter $\lambda > 0$, that is $Q_X = \text{Po}(\lambda)$. The resulting transition for this extended case

is given by

$$p_t(x_0, x_t) = \frac{e^{-\lambda\xi_t}\,(\lambda\xi_t)^{x_t}\,\xi_t^{x_0}}{x_t!}\,{}_2F_0\left(-x_0, -x_t; \frac{1-\xi_t}{\lambda\xi_t^2}\right)$$

where ξ can be also extended to $\xi_t = 1 - e^{-\alpha t}$ with $\alpha > 0$. All the results from Example 3.3 also follow in this case. In particular the infinitesimal generator is given by

$$q_{ij} = \begin{cases} -\alpha(i+\lambda), & i = j \\ \alpha\lambda, & j = i+1 \\ \alpha\, i, & j = i-1 \\ 0, & \text{otherwise.} \end{cases}$$

This also defines a birth and death process.

Thus, one question to ask is:

Which transition, or equivalently which $F_{Y|X}$, should we use?

The italicized question does not necessarily have a closed answer. For instance, in the Example 3.2, the situation might suggest to estimate the parameter s (which can be done, for example, using EM algorithm based on the augmented likelihood). However, if instead of $\mathrm{N}(x, s)$ we choose $\mathrm{St}(0, x, s)$ or some other distribution on the real line (possibly depending on other parameters) the choice turns out to be too wide.

In order to simplify the above issue, Pitt et al. [78] took the approach of imposing the further condition of linear expectation on the underlying transition, to reduce,

$$\mathbb{E}[X_{t+1} \mid X_t = x] = (1-\rho)\mu + \rho x, \tag{3.35}$$

where $|\rho| < 1$ and $\mu = \mathbb{E}[X]$. Based on this approach, they achieved the construction of general models, such as models with stationary distributions belonging to the class of convolution closed exponential families. See also Joe [48]. We will come back to this point in Chapter 4.

A natural way of choosing $F_{Y_{t+1}|X_t=x_t}(\cdot \mid x_t)$ might be based on aspects such as:

- Interpretability of the event $Y \mid X$.
- Availability of conditional data $Y \mid X$.
- Tractability of the transition $X_{t+1} \mid X_t$.
- Identifiability of the transition $X_{t+1} \mid X_t$ as a known model of $\{X_t\}_{t=1}^{\infty}$.
- Imposing a specific functional form for any of the conditionals or for the transition.

Interpretability and/or availability of $Y \mid X$ clearly simplifies the situation, since the choice of $F_{Y|X}$ can be set relatively easily, for example $Y \mid X$ may represent volatility given certain price level "X" which clearly induces certain parametric form. The tractability of the transition $X_{t+1} \mid X_t$ might help in the procedure to generalize the models to continuous time, see Examples 3.2, 4.1 and Section 6.2. Identifiability provides us with the latent representation in context, which allows the use of conditionals (3.24) and (3.25) in estimation procedures and to establish stationary conditions for the known model. For example the Poisson-gamma based models introduced in Chapter 6 give a representation for some well-known diffusion processes. By imposing specific functional forms to the conditional distributions or to the transition probability it is possible to achieve the construction for specific problems, for instance the ARCH models presented in Chapter 4.

The discussion given so far suggests that in order to construct stationary models, the reversibility assumption plays an important role, in fact without the use of the "latent structure" (represented through the process $\{Y(t); t \in \mathcal{T}\}$ or intrinsically by the augmentation used in the joint $F_{X,Y}$) it is possible to construct models with the required stationary distribution. In other words, given a reversible transition it is possible to construct a stationary model with given invariant distribution and without the use of the latent structure. However, as we will see, the availability of the underlying latent structure in the construction at issue will allow us to depict the constructed model in a way that eases estimation procedures. See Section 3.7 below.

Here, it is worth mention the conceptual difference in our construction due to the generality of the parameter $t \in \mathcal{T}$. While in the MCMC methods, such as the Gibbs sampler method, this parameter always takes the value "$t = 1$", or better said, one-step

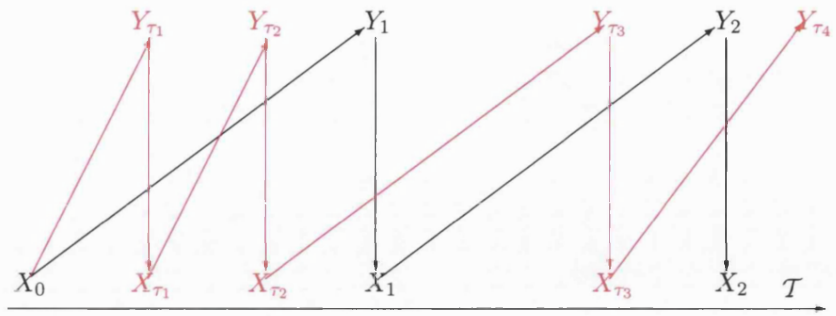


Figure 3.1: Dependence structure between latent and observed variables: discrete time (X_t, Y_t) and continuous time (X_{τ_t}, Y_{τ_t}). Data observed at $\tau_1, \tau_2 \ldots$ only need Y's at τ's.

transitions within the underlying state spaces, in the construction presented here $t \in \mathcal{T}$, with $\mathcal{T}$ not necessarily countable. For instance, in the illustration given above we can talk about the transition $x_s \to x_t$, where $s < t$, $s, t \in [0, \infty)$ and $x_s, x_t \in \mathbb{R}_+$ represent some time-space points on $\mathrm{E} \times \mathcal{T}$. This may imply, within the statistics context, that an infinite sample $x_{t_0}, x_{t_1}, \ldots$ is required. In Figure 3.1 an illustration that accounts for this discursion is given, sequences denoted with the subindex τ_i, $i = 1, 2, \ldots$ are in a continuous-time setting and sequences with subindex i are in a discrete-time setting.

Notice that with the Gibbs sampler type construction we are given $F_{Y|X}$ and Q_X, which implies the existence of the required $F_{X|Y}$. Equally, if we have the specification for Q_Y and $F_{X|Y}$ then we can find $F_{Y|X}$, however, in this case the stationary distribution Q_X is not given explicitly.

In contrast to the interleaving property corresponding to the two-component Gibbs sampler, where the independence structure only involves the immediate one-step (discrete) time-neighbors, in the general case (continuous time) the independence structure involves time-neighbors of an arbitrary time-distance.

Within the statistical framework, $y \in \mathrm{S}$ can be thought as the value of certain parameter in question and $F_{X|Y=y}$ as the model for the observations given a specific parameter value (or alternatively, after observing $X = x$, $F_{X|Y=y}$ as a function of y, which represents the likelihood). If inference about the parameter y is needed, then, from the Bayesian point of view, it can be done using the posterior probability. Another problem of interest in statistics is prediction, that is resolving the uncertainty of a "future" value

given a data set, in this case a set of observed x's. This can be done by constructing a transition (prediction) probability such as (3.22).

3.6 ALTERNATIVE CONSTRUCTION

The construction presented in Section 3.5 has the following alternative approach:

Let $\{X(t); t \in \mathcal{T}\}$ and $\{Y(t); t \in \mathcal{T}\}$ be two sequences of random variables $(\mathrm{E}, \mathcal{E})$ and $(\mathrm{S}, \mathcal{S})$ valued respectively and defined on a common probability space $(\Omega, \mathcal{A}, P)$. Assume $(\mathrm{E}, \mathcal{E})$ and $(\mathrm{S}, \mathcal{S})$ are Borel spaces. Suppose that $X_t \stackrel{\mathrm{d}}{=} Y$, $Y_t \stackrel{\mathrm{d}}{=} Y$, for all $t \in \mathcal{T}$ and denote $Q_X(B) = \mathrm{P}[X_t^{-1}(B)]$, $B \in \mathcal{E}$ and $Q_Y(C) = \mathrm{P}[Y_t^{-1}(C)]$, $C \in \mathcal{S}$ the unconditional distributions corresponding to the random variables X_t and Y_t. Furthermore assume the conditional independence structure (3.17) holds. As before, we regard $\{X(t); t \in \mathcal{T}\}$ the ***target process***. If $\mathcal{T}$ is countable and the corresponding spaces $(\mathrm{E}, \mathcal{E})$ and $(\mathrm{S}, \mathcal{S})$ are continuous (discrete), then models with independence structure such as (3.17), are known as general state-space models (hidden Markov models). See Elliot et al. [31] and Brockwell and Davis [20].

Notice that we are assuming the existence of an invariant measure Q_X, however, on the contrary of the construction presented in Section 3.5, here we start by specifying the conditionals $F_{X|Y}$ and $F_{Y|X}$ without full knowledge of the analytic form of Q_X. As before, in order to construct a valid transition probability that keeps Q_X invariant we can use the Gibbs sampler type update based on two conditional distributions such as (3.16). Starting with the two conditionals $F_{Y|X}$ and $F_{X|Y}$ without using the prior-posterior specification requires that the given conditionals come from the same joint distribution, so that the Gibbs sampler type update is valid.

When absolutely continuous densities do exist, one may speak about compatible conditional densities $f_{Y|X}(y \mid x)$ and $f_{X|Y}(x \mid y)$ for which Arnold and Press [5] give necessary and sufficient conditions for the existence of a joint distribution. Roughly speaking, in order to have a joint density $f_{X,Y}(x, y)$ with the above conditionals distributions, these must agree in their support, that is,

$$\{(x, y); f_{Y|X}(y \mid x) > 0\} = \{(x, y); f_{X|Y}(x \mid y) > 0\}$$

and there must exist functions $\nu(x)$ and $\eta(y)$ such that

$$\frac{f_{X|Y}(x \mid y)}{f_{Y|X}(y \mid x)} = \frac{\nu(x)}{\eta(y)}, \tag{3.36}$$

where $\int \nu(x) d\lambda_X(x) < \infty$ and $\int \eta(y) d\lambda_Y(y) < \infty$ are satisfied. Here, the marginal densities corresponding to the joint density $f_{X,Y}(x,y)$ are given by $q_X(x) \propto \nu(x)$ and $q_Y(y) \propto \eta(y)$. Hence, assuming the above conditions are satisfied we can express the joint density $f_{X,Y}(x,y)$ in terms of its compatible conditional densities as follows:

$$f_{X,Y}(x,y) = \frac{f_{Y|X}(y \mid x)}{\int f_{Y|X}(\xi \mid x)/f_{X|Y}(x \mid \xi)\, d\lambda(\xi)}. \tag{3.37}$$

In general, whichever the approach taken, either via the specification of the invariant distribution or the alternative construction, we have that the conditional independence structure (3.17) is satisfied.

Notice that in (3.21), $F_{X_s|Y_t=y}(\cdot) = F_{X_t|Y_t=y}(\cdot)$ due to the conditional independence assumptions (3.17). This implies that the time-shift, in the transition $\mathrm{P}_t(x,\cdot)$, is mainly driven by $F_{Y_t|X_0}$, while $F_{X_t|Y_t}$ only keeps the information of the updating time.

In the discrete-time setting, that is, when the target process is defined on a countable set $\mathcal{T}$, we focus on stationary models; hence a property to underline here is the finiteness of the invariant measure. See Chapter 4. In the continuous case, $\mathcal{T}$ uncountable, we will sometimes allow the constructed model to have no-finite invariant measures.

Transition probabilities constructed via (3.19) satisfy the Chapman-Kolmogorov equations when the time set $\mathcal{T}$ is assumed to be discrete. However, when the time is continuous Chapman-Kolmogorov equations are not straightforward and must be checked. For this reason we will treat both cases separately.

3.7 ESTIMATION OF LATENT STRUCTURE BASED MODELS

Let us assume that the conditional distributions, used in the underlying construction of a model for $\{X(t); t \in \mathcal{T}\}$, depend on certain parameter(s) denoted generically with $\theta \in \Theta$. If a tractable analytic expression for the transition density $p_t(x_0, x_t)$ is available, then we can directly maximize the likelihood for a given data set $\mathbf{x} = (x_{t_1}, \ldots, x_{t_T})$ for

$t_1 \leq t_2 \leq \cdots \leq t_T$, $t_1, \ldots, t_T \in \mathcal{T}$. That is, we need to compute $\max_\theta L_{\mathbf{x}}(\theta)$ where

$$L_{\mathbf{x}}(\theta) = q_X^\theta(x_{t_1}) \prod_{i=1}^{T-1} p_{(t_{i+1}-t_i)}(x_{t_i}, x_{t_{i+1}}). \tag{3.38}$$

The time-points $t_1, \ldots, t_T$ are not necessarily equally-spaced. In general, if we allow the sampling intervals to be random, then further assumptions on the distribution corresponding to this intervals should be made.

As an alternative, it may be easier to implement or computationally faster to consider the augmented likelihood and estimate the parameter(s) via the Expectation-Maximization (EM) algorithm. See Dempster et al. [26]. That is, since we can decompose the transition density on an integration over a latent variable, instead of carrying out the integration, which is not always easy analytically, we can consider the augmented likelihood

$$L_{\mathbf{x},\mathbf{y}}^{aug}(\theta) = q_X^\theta(x_{t_1}) \prod_{i=1}^{T-1} f_{X|Y}^\theta(x_{t_{i+1}} \mid y_{t_{i+1}}) \, f_{Y|X}^\theta(y_{t_{i+1}} \mid x_{t_i}). \tag{3.39}$$

It can be shown, see [26], that in order to maximize (3.38) we can compute iteratively a sequence $\theta_1, \ldots, \theta_j, \ldots$, converging to the maximum value of (3.38), with the following two steps

- E-step. For given data set $\mathbf{x}$ and current parameter value θ_j, compute the following expectation

$$Q(\theta \mid \theta_{(j)}, \mathbf{x}) = \mathbb{E}_{\theta_{(j)}} \left[\log L_{\mathbf{x},\mathbf{y}}^{aug}(\theta) \right], \tag{3.40}$$

  where the expectation $\mathbb{E}_{\theta_{(j)}}[\cdot]$ is taken with respect to $\mathrm{F}_{\mathbf{Y}|\mathbf{X}}^{\theta_{(j)}}$.

- M-step. Maximize $Q(\theta \mid \theta_{(j)}, \mathbf{x})$ in θ and define

$$\theta_{(j+1)} = \arg\max_\theta Q(\theta \mid \theta_{(j)}, \mathbf{x}). \tag{3.41}$$

The EM iterations satisfy

$$Q(\theta_{(j+1)} \mid \theta_{(j)}, \mathbf{x}) \geq Q(\theta_{(j)} \mid \theta_{(j)}, \mathbf{x}), \tag{3.42}$$

which implies that the sequence θ_j is always moving towards the maximum.
Notice that, in the case of our underlying construction, we can write

$$\mathrm{F}^{\theta}_{\mathbf{Y}|\mathbf{X}}(\mathbf{y} \mid \mathbf{x}) \propto \prod_{t=1}^{T-1} f^{\theta}_{X|Y}(x_{t_{i+1}} \mid y_{t_{i+1}})\, f^{\theta}_{Y|X}(y_{t_{i+1}} \mid x_{t_i}).$$

The main difficulty when implementing the EM algorithm lies on the E-step. Even when the conditional distribution $\mathrm{F}^{\theta_{(j)}}_{\mathbf{Y}|\mathbf{X}}$ can be computed analytically the integration required for the E-step may not be easy to evaluate. Under the assumption that a set $\mathbf{y}$ of latent random numbers is easily simulated from $\mathrm{F}^{\theta_{(j)}}_{\mathbf{Y}|\mathbf{X}}$, we can proceed from a Monte Carlo point of view and approximate Q as follows

$$\hat{Q}(\theta \mid \theta_{(j)}, \mathbf{x}) = \frac{1}{m} \sum_{k=1}^{m} \log(L^{aug}_{\mathbf{x},\mathbf{y}^{(k)}}(\theta)), \tag{3.43}$$

where $\mathbf{y}^{(k)} \sim \mathrm{F}^{\theta_{(j)}}_{\mathbf{Y}|\mathbf{X}}$. See Tanner and Wong [96] and Wei and Tanner [103]. However, for this Monte Carlo approach property (3.42) is not always satisfied due to the randomness of the augmented likelihood between iterations. Because of the independence structure underlying in this construction, we can simulate each component of $\mathbf{y}^{(k)}$ individually. That is, for a given k we can simulate $y^{(k)}$ by sampling individually each $y_{t_{i+1}}$ from a distribution with density

$$f(y_{t_{i+1}} \mid x_{t_{i+1}}, x_{t_i}) \propto f^{\theta}_{X|Y}(x_{t_{i+1}} \mid y_{t_{i+1}})\, f^{\theta}_{Y|X}(y_{t_{i+1}} \mid x_{t_i}) \tag{3.44}$$

for $i = 1, \ldots, T-1$.

Using the same functions as with the EM algorithm we can estimate the parameter(s), with a Bayesian point of view. Assume that (3.38) and (3.39) are integrable as functions of θ. Hence, we can consider the normalized likelihoods $L^{*}_{\mathbf{x}}(\theta)$ and $L^{*}_{\mathbf{x},\mathbf{y}}(\theta)$ respectively and implement the following Gibbs sampler:

$$\begin{aligned} \theta_{(j)} &\sim L^{*}_{\mathbf{x},\mathbf{y}^{(j-1)}}(\theta) \\ \mathbf{y}^{(j)} &\sim \mathrm{F}^{\theta_{(j)}}_{\mathbf{Y}|\mathbf{X}}(\mathbf{y} \mid \mathbf{x}). \end{aligned} \tag{3.45}$$

Notice that the sequence $\theta_{(j)}$ obtained from the Gibbs sampler (3.45) converges to a random variable with density $L^{*}_{\mathbf{x}}(\theta)$ as $j \to \infty$. See Casella and Berger [21].

An approximation to the normalized likelihood can be given by

$$\hat{L}^*_{\mathbf{x}}(\theta) = \frac{1}{m}\sum_{k=1}^{m} L^*_{\mathbf{x},\mathbf{y}^{(k)}}(\theta).$$

To sum up, we can consider, at least, four schemes:

1. ***MLE:*** Maximization of (3.38).
2. ***EM:*** EM algorithm, with E-step (3.40) and M-step (3.41).
3. ***MCEM:*** Monte Carlo EM (MCEM) algorithm, using approximation (3.43).
4. ***GSEM:*** The Gibbs sampler described in (3.45).

Further methods may be available due to the existence of the latent decomposition. In the remaining part of this thesis we will give some illustrations of the first three schemes described above.

3.8 SUMMARY

In Chapters 4 and 6 we will mainly use the Gibbs sampler type construction, described in Section 3.5, by distinguishing between the following two scenarios:

Discrete-time

Given that the invariant distribution Q_X belongs to a known parametric family, we will construct stationary time-homogeneous discrete-time Markov models $\mathrm{X} = \{X_t\}_{t=1}^{\infty}$ with such invariant distribution via the Gibbs sampler type construction. We will do this by suitable specifying a parametric family for the conditional distribution $F_{Y|X}$, required in the construction. As we have seen in page 48 such specification can be based on different aspects. In particular, this choice can be simplified by selecting a family for $F_{Y|X}$ such that a measurable function $h(\cdot)$ of the process $\mathrm{X} = \{X_t\}_{t=1}^{\infty}$ has the following linear expectation

$$\mathbb{E}[h(X_{t+1}) \mid X_t = x] = \rho\mu + (1-\rho)\, h(x),$$

where $0 < \rho < 1$ and μ is a constant. Using the Gibbs type update, the one-step transition density is given by

$$p(x_t, x_{t+1}) = \int f_{X|Y}(x_t \mid y)\, f_{Y|X}(y \mid x_{t+1})\, \lambda_Y(dy),$$

where $\lambda_Y(\cdot)$ is a σ-finite measure on the space of latent variables Y. With this transition density it is possible to construct the n-step transition probabilities satisfying the Chapman-Kolmogorov equation; see equation (4.3) in the next chapter.

An important point to underline here is that even in cases where the transition densities are not clearly available, the Gibbs sampler type construction is valid. Then the transition mechanism has the latent decomposition which in turn is useful in issues like estimation and simulation of the model.

Continuous-time

In this case we also assume that a known parametric family corresponding to the invariant distribution Q_X has been given. Hence, the objective here is to construct continuous-time homogeneous stationary Markov processes $\mathrm{X} = \{X_t; t \in \mathbb{R}_+\}$ with this invariant behavior. As in the discrete-time case, here we will also need to assume a parametric family for the conditional distribution $F_{Y|X}$. The difficulty in this case arises in that such conditional distribution is time-dependent, hence suitable ways to include this time information such that the Chapman-Kolmogorov equations are satisfied must be found. In particular, we will do this by allowing the parameters characterizing the family $F_{Y|X}$ to depend on time, namely $F^{\theta_t}_{Y|X}$ and then find suitable forms for θ_t such that the Chapman-Kolmogorov equations are satisfied.

It is worth mentioning that part of the objective in both scenarios is to recognize the constructed models as well-known models in the literature. This with the aim of using the latent decomposition in the estimation and simulation of the model.

CHAPTER 4

LATENT STRUCTURE BASED MODELS: DISCRETE TIME

In Chapter 3 the construction was presented for a general time-index set $\mathcal{T}$. Here, we confine ourselves to discrete-time models, in particular some generalizations from some of the models presented in Pitt and Walker [79] are studied. Section 4.1 provides with a more specific introduction within the discrete-time setting. In Section 4.2 definitions and some properties of generalized inverse Gaussian and generalized hyperbolic distributions are given. This definitions and properties are used to construct some stationary ARCH(1) models presented in Section 4.3. A generalization to higher lag-dependence stationary processes together with the corresponding generalizations of the ARCH models is studied in Section 4.5. For the introduced discrete models three estimation methods are analyzed in Sections 4.8, 4.9 and 4.10, the MLE, the EM and the MCEM respectively. Finally, some illustrations with real financial data are presented in Section 4.11.

4.1 INTRODUCTION

In Chapter 3 we gave a method to construct versions of reversible Markov processes for a given invariant measure. The construction simplifies considerably when the index set $\mathcal{T}$ of the process is taken to be countable, that is when the Chapman-Kolmogorov equations are intrinsically satisfied. This resembles much of the construction presented in Example 3.2.

As we saw before in Example 3.2, given a stationary distribution there are many models with the property of having the same stationary distribution, it is enough to change the form for the required conditional $F_{Y|X}$ to get a different model. Remember that an important component in the construction is the parametric family to which the conditional distribution $F_{Y|X}$ belongs (which in turn can be given by imposing forms for $F_{X|Y}$ and Q_Y).

In this chapter we generalize ARCH-type models by looking for conditional distributions $F_{Y|X}$ such that a specific functional relation for the transition probability holds. This generalization will be done along the same lines treated in Pitt and Walker [79]. Before starting the construction of such models we re-state the general idea within the discrete-time setting.

In this chapter, we assume that the *general setting with target process* X described in Chapter 3 holds in discrete time $\mathcal{T}$. For simplicity we assume $\mathcal{T} = \mathbb{N}$. Therefore, considering either the approach taken in Section 3.5 or the one in Section 3.6, we have two process $\mathrm{X} = \{X_t\}_{t=1}^{\infty}$, $\mathrm{Y} = \{Y_t\}_{t=1}^{\infty}$ satisfying dependence structure (3.17), which in this case can be written as presented in equations (3.24) and (3.25).

As before, let us assume that $F_{X_{t+1}|Y_{t+1}=y} \ll \lambda_X$ on the space E and $F_{Y_{t+1}|X_t=x} \ll \lambda_Y$ on the space S. Let $f_{X_{t+1}|Y_{t+1}}(x \mid y)$ and $f_{Y_{t+1}|X_t}(y \mid x)$ be their corresponding frequency functions.

With the above setting, the conditional independence structure (3.17) and the fact that $\mathcal{T}$ is discrete, we can proceed to construct a one-step Markov transition probability as follows

$$\mathrm{P}(x, B) = \int_B \int_S f_{X_{t+1}|Y_{t+1}}(\xi \mid y)\, f_{Y_{t+1}|X_t}(y \mid x) \lambda_Y(dy) \lambda_X(d\xi), \tag{4.1}$$

for $B \in \mathcal{E}$. With the one-step transition probability (4.1) it is possible to construct the n-step transition probabilities which satisfy the Chapman-Kolmogorov equations. Given the one-step transition probabilities (4.1) the n-step transition probability is computed as follows: Set $\mathrm{P}^0(x, B) = \delta_x(B)$ and for $n \geq 1$ define inductively

$$\mathrm{P}^n(x, B) = \int_E \mathrm{P}(x, dy)\, \mathrm{P}^{n-1}(y, B), \tag{4.2}$$

where $x \in \mathrm{E}$ and $B \in \mathcal{E}$. Using expression (4.2), the Chapman-Kolmogorov equations can be written as

$$\mathrm{P}^n(x, B) = \int_E \mathrm{P}^m(x, dy)\, \mathrm{P}^{n-m}(y, B), \quad 0 \leq m \leq n. \tag{4.3}$$

In general, if a joint sequence $\{Z_t\}_{t=1}^{\infty}$ is defined by $Z_t = (X_t, Y_t)$, then a one-step Markov transition probability on $\mathrm{E} \times \mathrm{S}$ can be constructed as follows

$$\mathrm{P}(z, A) = \int_A f_{X_{t+1}|Y_{t+1}}(x' \mid y')\, f_{Y_{t+1}|X_t}(y' \mid x)\, \lambda(dx' \times dy'), \tag{4.4}$$

where $z = (x, y) \in \mathrm{E} \times \mathrm{S}$, $A \in \mathcal{E} \otimes \mathcal{S}$ and $\lambda(\cdot)$ is a σ-finite measure on $\mathrm{E} \times \mathrm{S}$. Notice the independence of the above transition with respect to the initial state y. The one-step transition probabilities for the marginal process $\{Y_t\}_{t=1}^{\infty}$ can be equally found as the marginals of (4.4). It is well known within the Gibbs sampler literature, that although the underlying construction ensures that the processes X and Y are reversible, this property does not necessarily holds for the joint process defined through (4.4).

Example 4.1. (Example 3.2 continued)
In Example 3.2 we presented a process with invariant distribution $Q_X = N(0, 1)$ obtained by specifying $F_{Y|X=x} = \mathrm{N}(x, s)$, $s > 0$. This led to a one-step transition probability given by

$$\mathrm{P}(x, \cdot) = \mathrm{N}\left(\frac{x}{s+1}, \frac{(2+s)s}{(s+1)^2}\right). \tag{4.5}$$

The above model can be written as

$$X_{t+1} = \frac{X_t}{s+1} + \sqrt{\frac{(2+s)s}{(s+1)^2}}\, \varepsilon_{t+1}, \quad \varepsilon_{t+1} \sim \mathrm{N}(0, 1)$$

which can be identified as a Gaussian AR(1) model.

We are interested in computing the n-step transition probabilities, which will be used for a generalization to the continuous time version of the model.

First let us notice that, in general, if we are given a transition probability $\mathrm{P}(x, \cdot) = \mathrm{N}(xa, b)$, $b > 0$ then $\mathrm{P}^2(x, \cdot) = \mathrm{N}(xa^2, b(a^2 + 1))$. Hence using (4.3), we can use P^2 to

get P^3, which is given by $N(xa^3, b(a^4 + a^2 + 1))$. Working iteratively, we obtain

$$P^n(x, \cdot) = N\left(xa^n, b\sum_{i=1}^{n} a^{2(i-1)}\right). \tag{4.6}$$

Accordingly, we substitute

$$a = \frac{1}{s+1} \quad \text{and} \quad b = \frac{(2+s)s}{(s+1)^2}$$

in transition probabilities (4.5). Notice that for $s > 0$ we have that $1/(s+1) < 1$, implying that the geometric sum corresponding to the variance term in (4.6) is convergent, explicitly

$$\sum_{i=1}^{n}(s+1)^{-2(i-1)} = \frac{(s+1)^2\left(1-s^2+2s+1\right)^{-n}}{s\,(s+2)}.$$

Therefore, the n-step transition probability, is given by

$$P^n(x, \cdot) = N\left(\frac{x}{(s+1)^n}, 1 - \frac{1}{(s+1)^{2n}}\right). \tag{4.7}$$

Here we see that, $P^n(x, B) \to Q_X(B)$, for all $B \in \mathcal{E}$, as $n \to \infty$, hence Q_X is not only the invariant distribution but also the limiting distribution. Figure 4.1 illustrates this fact. See Section 6.2 for a generalization of this example to continuous time. o

In Example 4.1, the n-step transition probability was easily obtained. The integral in (4.2) was analytically easy to evaluate, due to the tractability of the normal distribution. However, in many cases such integrals are more complicated having to be computed numerically.

The idea of using reversibility as a constructive tool of stationary models was introduced in Pitt et al. [78]. In their approach, an extra moment condition for the transition mechanism was assumed by imposing the linear expectation (3.35). As stated in [78], this linear assumption still results in a wide choice for the distributional family to which $F_{Y|X}$ belongs. However, Pitt et al. [78] showed that the general approach taken in Jørgensen and Song [50] can be seen as a particular case of the reversibility-based construction.

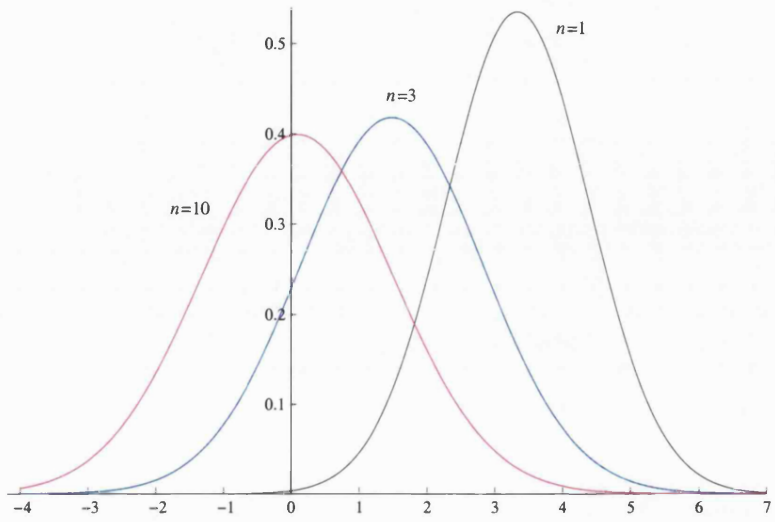


Figure 4.1: n-step transition density for the normal model in Example 4.1. Starting point $x = 5$ and $s = 0.5$.

To recap, let us assume that we want to construct a stationary Markov process $\{X_t\}_{t=1}^{\infty}$ with invariant distribution $\mathrm{ED}^*(\theta;\tau)$, where $\tau = \tau_1 + \tau_2$. Assume $\tau_1 = \rho\lambda$, $\tau_2 = (1-\rho)\lambda$ and $0 < \rho < 1$ and that $F_{Y|X}(\cdot \mid x) = G(\rho\lambda, (1-\rho)\lambda, x)$ (see equations (3.6) and (3.8)). After an application of Bayes theorem we see that

$$f_{X|Y}(x \mid y) = c(x-y;(1-\rho)\lambda)\, \exp\left\{(x-y)\theta - (1-\rho)\lambda k(\theta)\right\}.$$

Notice that, with the above assumptions, we have

$$\mathbb{E}[Y] = \tau_1 k'(\theta) = \rho\lambda k'(\theta) = \rho\, \mathbb{E}[X]$$

and therefore $\mathbb{E}[\mathbb{E}[Y \mid X]] = \rho\, \mathbb{E}[X]$ leading to

$$\mathbb{E}[Y \mid X = x] = \rho x. \tag{4.8}$$

Now, since

$$F_{X|Y}(x \mid y) = F_{Y_2}(x-y),$$

where $Y_2 \sim \mathrm{ED}^*(\theta;\tau_2)$, then

$$\mathbb{E}[X \mid Y] = Y + \tau_2 k'(\theta),$$

resulting in

$$\begin{aligned}
\mathbb{E}[X_{t+1} \mid X_t = x] &= \mathbb{E}[\mathbb{E}[X_{t+1} \mid Y_{t+1}] \mid X_t = x] \\
&= \mathbb{E}[Y_{t+1} + \tau_2 k'(\theta) \mid X_t = x] \\
&= \mathbb{E}[Y_{t+1} \mid X_t = x] + \tau_2 k'(\theta) \\
&= x\rho + (1-\rho)\,\mathbb{E}[X]. \qquad (4.9)
\end{aligned}$$

In equality (4.9) we have used the conditional expectation (4.8). Even though in this case the underlying transition probability does not have a tractable analytic expression (in the general case), the constructed process has the required invariant distribution and satisfies the linearity form (4.9) imposed in Pitt et al. [78]. Furthermore, when direct MLE methods are not available, the availability of the latent structure gives the required decomposition to directly apply the EM estimation method described in Section 3.7.

The autocorrelation of a time-homogenous stationary Markov model $\{X_t\}_{t=1}^{\infty}$ with the linear structure (4.9) can be computed by considering the following recursive mechanism:

$$\begin{aligned}
\mathbb{E}[X_{t+1} \mid X_t = x] &= \rho x + (1-\rho)\mu \\
\mathbb{E}[X_{t+2} \mid X_t = x] &= \mathbb{E}\left[\mathbb{E}[X_{t+2} \mid X_{t+1}, X_t] \mid X_t = x\right] = \mathbb{E}\left[\mathbb{E}[X_{t+2} \mid X_{t+1}] \mid X_t = x\right] \\
&= \mathbb{E}\left[\rho X_{t+1} + (1-\rho)\mu \mid X_t = x\right] = \rho\,\mathbb{E}[X_{t+1} \mid X_t = x] + (1-\rho)\mu \\
&= \rho^2 x + (1-\rho^2)\mu \\
&\vdots \quad \vdots \\
\mathbb{E}[X_{t+h} \mid X_t = x] &= \mathbb{E}\left[\mathbb{E}\left[\cdots \mathbb{E}\left[X_{t+h} \mid X_{t+h-1}\right] \mid X_{t+h-2} \cdots\right] \mid X_t = x\right] \\
&= \rho^h x + (1-\rho^h)\mu.
\end{aligned}$$

Where $\mu = \mathbb{E}[X_t]$, that is the mean of the stationary distribution. Now

$$\mathbb{E}[X_{t+h}X_t] = \mathbb{E}[X_t \mathbb{E}\left[X_{t+h} \mid X_t\right]] = \rho^h\,\mathbb{E}[X_t^2] + (1-\rho^h)\mu^2.$$

Hence, using the stationarity of the model, the autocorrelation is given by

$$\text{Corr}(X_{t+h}, X_t) = \frac{\mathbb{E}[X_{t+h}X_t] - \mu^2}{\mathbb{E}[X_t^2] - \mu^2} = \rho^h. \tag{4.10}$$

This clearly provides with an extra parameter that could be used to fit a required autocorrelation, rather than only focusing on the stationary distribution.

4.2 GENERALIZED HYPERBOLIC DISTRIBUTIONS

In many modelling contexts, such as econometrics, more flexible distributions, in particular those allowing heavier tails and skewness, are typically of interest. For example, it is well known that the log-returns of financial assets are better fitted with distributions having heavier tails than the normal distribution. See, Mandelbrot [69], Fama [33], Mittnik and Rachev [75] and more recently Eberlein [29].

Recently it has been widely recognized, from a distributional perspective, that generalized hyperbolic (GH) distributions offer an adequate fit for modelling the log-returns of financial instruments. Therefore, with this application in mind, we devote this section to define and study some properties of this families.

Generalized hyperbolic distributions were first introduced by Barndorff-Nielsen [7] in order to model grain size distributions of wind blown sands. Nowadays GH distributions play an important role in financial applications, see for instance Barndorff-Nielsen and Shephard [13].

Definition 4.1. The random variable X has a ***generalized hyperbolic distribution*** *(GH)* if its density with respect to the Lebesgue measure is defined as

$$\begin{aligned} \text{GH}(x; \lambda, \alpha, \beta, \delta, \mu) &= a(\lambda, \alpha, \beta, \delta) \left\{\delta^2 + (x-\mu)^2\right\}^{(\lambda - \frac{1}{2})/2} \\ &\times K_{\lambda - 1/2}\left(\alpha\sqrt{\delta^2 + (x-\mu)^2}\right) \exp\left\{\beta(x-\mu)\right\} \end{aligned} \tag{4.11}$$

with

$$a(\lambda, \alpha, \beta, \delta) = \frac{\left(\alpha^2 - \beta^2\right)^{\lambda/2}}{\sqrt{2\pi}\alpha^{\lambda - \frac{1}{2}}\delta^\lambda K_\lambda\left(\delta\sqrt{\alpha^2 - \beta^2}\right)},$$

where $x \in \mathbb{R}$ and K_ν is the modified Bessel function[1] of the third kind with index ν. The above density depends on five parameters: $\alpha > 0$ as a shape parameter, β with $0 \leq |\beta| < \alpha$ determines the skewness, $\mu \in \mathbb{R}$ the location, $\delta > 0$ as a scaling factor and the parameter $\lambda \in \mathbb{R}$ which characterizes certain sub-classes and is related to the amount of mass in the tails.

Generalized hyperbolic distributions can be seen as a mean-variance mixture of normal distributions where the mixing distribution is ***generalized inverse Gaussian*** (GIG), that is, a distribution with Lebesgue density given by

$$\mathrm{GIG}\,(x;\lambda,\delta,\gamma) = \frac{\left(\frac{\gamma}{\delta}\right)^{\lambda/2}}{2K_\lambda(\sqrt{\delta\gamma})}\, x^{\lambda-1} \exp\left\{-\frac{1}{2}\left(\delta x^{-1} + \gamma x\right)\right\}, \tag{4.12}$$

where $\lambda \in \mathbb{R}$, $(\delta,\gamma) \in \Theta_\delta$ and

$$\theta_\lambda = \begin{cases} \delta \geq 0, \gamma > 0 & \text{if } \lambda > 0 \\ \delta > 0, \gamma > 0, & \text{if } \lambda = 0 \\ \delta > 0, \gamma \geq 0, & \text{if } \lambda < 0. \end{cases} \tag{4.13}$$

The cases $\delta = 0$ and $\gamma = 0$ are interpreted as the limiting cases. Figure 4.2 shows how GH distributions vary as the parameters change.

The GH distribution can be obtain as

$$\mathrm{GH}(x;\lambda,\alpha,\beta,\delta,\mu) = \int_0^\infty \mathrm{N}(x;\mu+\beta y, y)\,\mathrm{GIG}(y;\lambda,\delta^2,\alpha^2-\beta^2)\,dy. \tag{4.14}$$

One of the main advantages of using generalized hyperbolic distributions is their generality embracing many special cases. This generality is inherited from the representation (4.14) and therefore from the generality of GIG distributions. Many of the special cases are clarified via the Laplace transform. In the case of the GIG distribution, the Laplace transform is given by

$$\begin{aligned} \mathcal{L}_{GIG}(\theta) &= \int_0^\infty e^{\theta x}\,\mathrm{GIG}(x;\lambda,\delta,\gamma)dx \\ &= \left(\frac{\gamma}{\gamma-2\theta}\right)^{\frac{\lambda}{2}} \frac{K_\lambda(\sqrt{\delta(\gamma-2\theta)})}{K_\lambda(\sqrt{\delta\gamma})}, \quad 2\theta < \gamma, \end{aligned} \tag{4.15}$$

[1]Modified Bessel functions $K_\nu(z)$ can be defined as the solutions to the differential equation $zw'' + zw' - (z^2+\nu^2)w = 0$, see Abramowitz and Stegun [2].

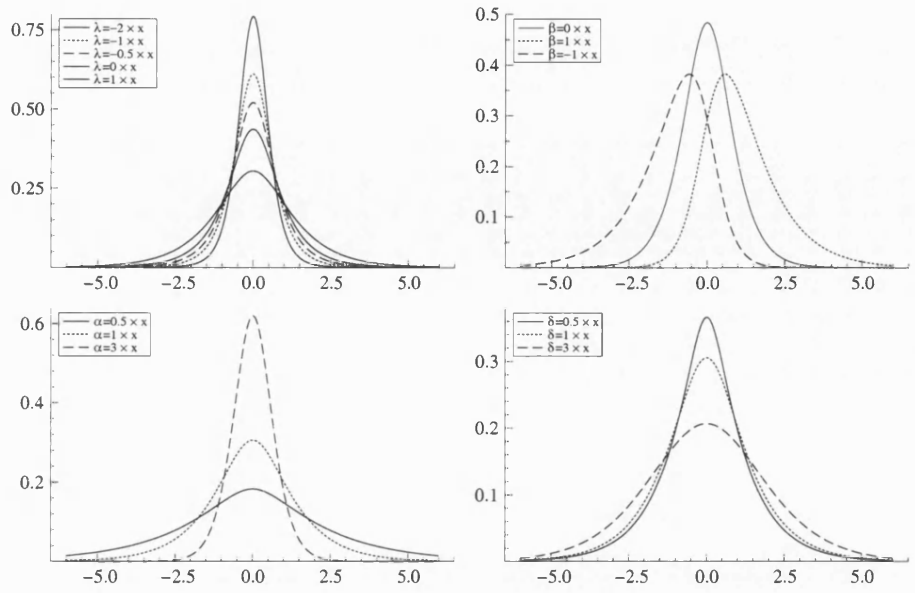


Figure 4.2: Densities of a generalized hyperbolic distribution as parameters vary.

which in turn, using (4.14), implies

$$\begin{aligned}\mathcal{L}_{GH}(\theta) &= e^{\mu\theta}\mathcal{L}_{GIG(\lambda,\delta^2,\alpha^2-\beta^2)}\left(\frac{\theta^2}{2}+\beta\theta\right)\\ &= e^{\mu\theta}\left\{\frac{\alpha^2-\beta^2}{\alpha^2-(\beta+\theta)^2}\right\}^{\frac{\lambda}{2}}\frac{K_\lambda(\delta\sqrt{\alpha^2-(\beta+\theta)^2})}{K_\lambda(\delta\sqrt{\alpha^2-\beta^2})}, \quad |\beta+\theta|<\alpha.\end{aligned}$$

By means of the above expressions and some basic properties of the modified Bessel function K_ν (Table 4.1) it is possible to characterize some well-known families.

Using the asymptotic expansion 1 in Table 4.1 we can see that if $\lambda > 0$ and $\delta \downarrow 0$, then density (4.12) reduces to

$$\mathrm{GIG}(x;\lambda,0,\gamma) = \frac{\left(\frac{\gamma}{2}\right)^\lambda}{\Gamma(\lambda)}x^{\lambda-1}e^{-\frac{\gamma}{2}x} = \mathrm{Ga}\left(x;\lambda,\frac{\gamma}{2}\right),$$

characterizing as a special case the gamma distribution. Analogously, when $\lambda < 0$, $\gamma \downarrow 0$ and using the asymptotic expansion 2 in Table 4.1 we get

$$\mathrm{GIG}(x;\lambda,\delta,0) = \frac{\left(\frac{2}{\delta}\right)^\lambda}{\Gamma(-\lambda)}x^{\lambda-1}e^{-\frac{\delta}{2x}} = \mathrm{Iga}\left(x;-\lambda,\frac{\delta}{2}\right), \tag{4.16}$$

Modified Bessel functions	
Properties	Asymptotic expansions as $x \downarrow 0$
1.- $K_{-\nu}(x) = K_\nu(x)$	1.- $K_\nu(x) \sim \frac{1}{2}\Gamma(\nu)\left(\frac{x}{2}\right)^{-\nu}, \quad \nu > 0$
2.- $K_{1/2}(x) = \sqrt{\frac{\pi}{2x}}e^{-x}$	2.- $K_\nu(x) \sim \frac{1}{2}\Gamma(-\nu)\left(\frac{x}{2}\right)^{\nu}, \quad \nu < 0$
3.- $K_{\nu+\varepsilon}(x) > K_\nu(x), \quad \nu, \varepsilon, x > 0$	3.- $K_0(x) \sim -\log(x)$
4.- $K_{\nu+1}(x) = \frac{2\lambda}{x}K_\nu(x) + K_{\nu-1}(x)$	
Integral representation	Asymptotic expansion as $x \uparrow \infty$
$K_\nu(x) = \frac{1}{2}\int_0^\infty y^{\nu-1}\exp\left(-\frac{x}{2}(y+y^{-1})\right)dy$	$K_\nu(x) \sim \sqrt{\frac{\pi}{2x}}e^{-x}$
Derivatives $\partial/\partial x$	
1.- $K_0'(x) = -K_1(x)$	4.- $(\log K_\nu(x))' = \frac{\nu}{x} - R_\nu(x)$
2.- $K_\nu'(x) = -\frac{1}{2}(K_{\nu+1}(x) + K_{\nu-1}(x))$	5.- $(\log K_\nu(x))'' = S_\nu(x) - \frac{R_\nu(x)}{x} - \frac{\nu}{x^2}$
3.- $K_\nu'(x) = \frac{\lambda}{x}K_\nu(x) - K_{\nu+1}(x)$	
where $R_\nu(x) := \frac{K_{\nu+1}(x)}{K_\nu(x)}$, $x > 0$ and	$S_\nu(x) := \frac{K_{\nu+2}(x)K_\nu(x) - K_{\nu+1}^2(x)}{K_\nu^2(x)}$, $x > 0$

Table 4.1: Some relevant properties of the modified Bessel function. Reference: Abramowitz and Stegun [2] (Chapter 9) and Eberlein and Hammerstein [30].

characterizing the inverse gamma distribution. In addition, the parameter λ also characterizes some important families, for instance when $\lambda = -1/2$, (4.12) characterizes the inverse Gaussian family, $\mathrm{IG}(x; \delta, \gamma)$, with density

$$\mathrm{GIG}\left(x; -1/2, \delta^2, \gamma^2\right) = \frac{\delta}{\sqrt{2\pi}}\, e^{\delta\gamma}\, x^{-3/2} \exp\left\{-\frac{1}{2}(\delta^2 x^{-1} + \gamma^2 x)\right\} = \mathrm{IG}(x; \delta, \gamma).$$

Particular cases of the GIG distribution lead to particular cases of the mixture (4.14), that is particular cases of generalized hyperbolic distributions. For example, the normal inverse Gaussian (NIG) distribution can be obtained as

$$\begin{aligned}
\mathrm{GH}(x; -1/2, \alpha, \beta, \delta, \mu) &= \int_0^\infty \mathrm{N}(x; \mu + \beta y, y)\, \mathrm{IG}(y; \delta, \sqrt{\alpha^2 - \beta^2})dy \\
&= \frac{\alpha\delta}{\pi} \exp\left\{\delta\sqrt{\alpha^2 - \beta^2} + \beta(x-\mu)\right\} \frac{K_1\left(\alpha\sqrt{\delta^2 + (x-\mu)^2}\right)}{\sqrt{\delta^2 + (x-\mu)^2}} \\
&= \mathrm{NIG}(x; \alpha, \beta, \delta, \mu).
\end{aligned}$$

Many other cases are encompassed, we refer to Jørgensen [49] and Eberlein and Hammerstein [30] for further particular cases and results on GIG and GH distributions.

Theorem 4.1. (***Moments***) If $X \sim \mathrm{GH}(\lambda, \alpha, \beta, \delta, \mu)$ then the n-th non-central moments $\mathrm{M}_{GH}^{(n)}$ are characterized through the following recursion formula

$$\mathrm{M}_{GH}^{(n)} = \sum_{i=0}^{n} \binom{n}{i} \mu^{n-i}\, \mathrm{M}_{GIG}(i, \beta, \omega, \eta), \quad n = 1, 2, \ldots$$

where $\mathrm{M}_{GIG}(0,\beta,\omega,\eta)=1$ and

$$\mathrm{M}_{GIG}(n,\beta,\omega,\eta)=\begin{cases}\sum_{i=1}^{r}\frac{(2r-1)!\beta^{2i-1}}{(2i-1)!(r-i)!2^{r-i}}\frac{K_{\lambda+r+i-1}(\omega)\eta^{r+i-1}}{K_\lambda(\omega)}, & n=2r-1\\ \sum_{i=0}^{r}\frac{(2r)!\beta^{2i}}{(2i)!(r-i)!2^{r-i}}\frac{K_{\lambda+r+i}(\omega)\eta^{i+r}}{K_\lambda(\omega)}, & n=2r\end{cases}$$

with $\omega=\delta\sqrt{\alpha^2-\beta^2}$ and $\eta=\delta/\sqrt{\alpha^2-\beta^2}$.

Proof. First, let us notice that $\mathrm{M}_{GH}^{(n)}=\mathcal{L}_{GH}^{(n)}(0)$, where $\mathcal{L}_{GH}(\theta)=e^{\mu\theta}\mathcal{L}_{GIG}(g(\theta))$ and $g(\theta)=\theta^2/2+\beta\theta$. Notice that $g'(\theta)=\theta+\beta$, $g''(\theta)=1$ and $g^{(n)}(\theta)=0$ for $n\geq 3$. Using a simple binomial induction, we get

$$\begin{aligned}\mathcal{L}_{GH}^{(1)}(\theta) &= e^{\mu\theta}\left[\mu\mathcal{L}_{GIG}(g(\theta))+\left(\frac{\partial}{\partial\theta}\right)\mathcal{L}_{GIG}(g(\theta))\right]\\ \mathcal{L}_{GH}^{(2)}(\theta) &= e^{\mu\theta}\left[\mu^2\mathcal{L}_{GIG}(g(\theta))+2\mu\left(\frac{\partial}{\partial\theta}\right)\mathcal{L}_{GIG}(g(\theta))+\left(\frac{\partial}{\partial\theta}\right)^2\mathcal{L}_{GIG}(g(\theta))\right]\\ &\vdots\\ \mathcal{L}_{GH}^{n}(\theta) &= e^{\mu\theta}\sum_{i=0}^{n}\binom{n}{i}\mu^{n-i}\left(\frac{\partial}{\partial\theta}\right)^i\mathcal{L}_{GIG}(g(\theta)).\end{aligned}\tag{4.17}$$

In order to compute the differential operator in equation (4.17) notice that the moments corresponding to a normal distribution $\mathrm{N}(\beta y,y)$ are given by

$$\mu_n'(y)=\begin{cases}\sum_{i=1}^{r}\frac{(2r-1)!\beta^{2i-1}y^{i+r-1}}{(2i-1)!(r-i)!2^{r-i}}, & n=2r-1\\ \sum_{i=0}^{r}\frac{(2r)!\beta^{2i}y^{i+r}}{(2i)!(r-i)!2^{r-i}}, & n=2r.\end{cases}\tag{4.18}$$

See Bain [6]. Therefore, we can define the operator $\mathrm{D}_n y$ by replacing in (4.18) all the powers of y with the corresponding differentiation order, that is we replace y^i with $y^{(i)}$, and obtain

$$\left(\frac{\partial}{\partial\theta}\right)^n\mathcal{L}_{GIG}(g(\theta))=\mathrm{D}_n\mathcal{L}_{GIG}(\theta).\tag{4.19}$$

Hence, by noticing, from (4.15), that

$$\mathcal{L}_{GIG}^{(n)}(0)=\frac{K_{\lambda+n}(\omega)}{K_\lambda(\omega)}\eta^n,\tag{4.20}$$

and evaluating $\theta=0$ in (4.17) the stated result follows. □

By applying Theorem 4.1 we can compute some useful moments, for example, if $X \sim \mathrm{GH}(x; \lambda, \alpha, \beta, \delta, \mu)$ then

$$\mathbb{E}[X] = \mu + \beta\eta \frac{K_{\lambda+1}(\omega)}{K_\lambda(\omega)} \tag{4.21}$$

$$\mathbb{E}[X^2] = \mu^2 + \frac{K_{\lambda+1}(\omega)}{K_\lambda(\omega)} \eta(2\mu\beta + 1) + \eta^2\beta^2 \frac{K_{\lambda+2}(\omega)}{K_\lambda(\omega)} \tag{4.22}$$

$$\begin{aligned}\mathbb{E}[X^3] &= \mu^3 + \frac{K_{\lambda+1}(\omega)\eta}{K_\lambda(\omega)} \left(3\mu^2\beta + 3\mu\right) + \frac{K_{\lambda+2}(\omega)\eta^2}{K_\lambda(\omega)} \left(3\mu\beta^2 + 3\beta\right) \\ &+ \frac{\beta^3\eta^3 K_{\lambda+3}(\omega)}{K_\lambda(\omega)}\end{aligned} \tag{4.23}$$

$$\begin{aligned}\mathbb{E}[X^4] &= \mu^4 + \frac{K_{\lambda+1}(\omega)\eta}{K_\lambda(\omega)} \left(4\mu^3\beta + 6\mu^2\right) + \frac{\eta^2 K_{\lambda+2}(\omega)}{K_\lambda(\omega)} \left(6\mu^2\beta^2 + 12\mu\beta + 3\right) \\ &+ \frac{K_{\lambda+3}(\omega)\eta^3}{K_\lambda(\omega)} \left(4\mu\beta^3 + 6\beta^2\right) + \frac{\beta^4\eta^4 K_{\lambda+4}(\omega)}{K_\lambda(\omega)}.\end{aligned} \tag{4.24}$$

Having introduced generalized inverse Gaussian distributions, we take the opportunity to illustrate, in the following example, an alternative way of constructing stationary discrete-time Markov models.

Example 4.2. (Jørgensen [49]) Using a similar approach as the one presented in Joe [48] and Jørgensen and Song [50] (see Section 3.2) it is possible to construct a stationary time series with GIG transition probabilities. Let us consider the re-parametrization $\omega = \sqrt{\delta\gamma}$ and $\eta = \sqrt{\delta/\gamma}$. Assume

$$X_i \overset{\text{ind}}{\sim} \mathrm{GIG}(\lambda_i, \delta_i, \gamma_i), \quad i = 1, 2$$

and consider the transformations $T = \sqrt{X_1 X_2}$ and $S = \sqrt{X_2/X_1}$. Therefore, we have the following joint density

$$\begin{aligned}f_{S,T}(s,t) &= \frac{\eta_1^{-\lambda_1}\eta_2^{-\lambda_2}}{2K_{\lambda_1}(\omega_1)K_{\lambda_2}(\omega_2)} t^{\lambda_1+\lambda_2-1} s^{\lambda_2-\lambda_1-1} \\ &\times \exp\left\{-\frac{1}{2}\left(s^{-1}(\gamma_1 t + \delta_2 t^{-1}) + s(\delta_1 t^{-1} + \gamma_2 t)\right)\right\}\end{aligned} \tag{4.25}$$

with corresponding conditionals

$$f_{S|T}(s \mid t) = \mathrm{GIG}(s; \lambda_2 - \lambda_1, \gamma_1 t + \delta_2 t^{-1}, \delta_1 t^{-1} + \gamma_2 t) \tag{4.26}$$

$$f_{T|S}(t \mid s) = \mathrm{GIG}(t; \lambda_1 + \lambda_2, \delta_2 s^{-1} + \delta_1 s, \gamma_1 s^{-1} + \gamma_2 s). \tag{4.27}$$

If we fix $\delta_1 = \gamma_1 = \varpi$ and $\lambda_1 = 0$ then $f_{S|T} = f_{T|S}$ and at the same time $S \overset{\mathrm{d}}{=} T$, since GIG distributions are strictly positive. With this, it is possible to take one of the conditional densities (4.26) or (4.27) as the one-step transition density associated to a discrete-time Markov process. Hence, we can define

$$p(x_t, x_{t+1}) := \mathrm{GIG}(x_{t+1}; \lambda_2, x_t\varpi + \delta_2 x_t^{-1}, \varpi x_t^{-1} + \gamma_2 x_t). \tag{4.28}$$

Proposition 4.1. The invariant density for a discrete-time Markov process with the above one-step transition density is given by

$$q_X(u) = \frac{K_{\lambda_2}\left(\sqrt{\varpi u + \delta_2/u}\sqrt{\varpi/u + \gamma_2 u}\right)}{K_0(\varpi)\, K_{\lambda_2}(\omega_2)\, \eta_2^{\lambda_2}\, u^{1-\lambda_2}} \left(\frac{\varpi u + \delta_2 u^{-1}}{\varpi u^{-1} + \gamma_2 u}\right)^{\frac{\lambda_2}{2}}.$$

Proof. The invariant density corresponding for a Markov model with one-step transition density (4.28) is given by either $f_S(u)$ or $f_T(u)$, since we have chosen as the transition density one of the conditionals $f_{S|T}$ or $f_{T|S}$, which in turn, for $\delta_1 = \gamma_1 = \varpi$ and $\lambda_1 = 0$, are the same. Notice that for these parameter values, the re-parametrization η_1 is given by $\eta_1 = \sqrt{\delta_1/\gamma_1} = 1$. Hence, a direct marginalization of the joint density (4.25) leads to

$$\begin{aligned} q_X(u) &= \int_0^\infty f_{S,T}(u,t)\, dt \\ &= \frac{\eta_2^{-\lambda_2} u^{\lambda_2 - 1}}{2K_0(\varpi)K_{\lambda_2}(\omega_2)} \\ &\times \int_0^\infty t^{\lambda_2 - 1} \exp\left\{-\frac{1}{2}\left(u^{-1}(\varpi t + \delta_2 t^{-1}) + u(\varpi t^{-1} + \gamma_2 t)\right)\right\} dt \\ &= \frac{K_{\lambda_2}\left(\sqrt{\varpi u + \delta_2/u}\sqrt{\varpi/u + \gamma_2 u}\right)}{K_0(\varpi)\, K_{\lambda_2}(\omega_2)\, \eta_2^{\lambda_2}\, u^{1-\lambda_2}} \left(\frac{\varpi u + \delta_2 u^{-1}}{\varpi u^{-1} + \gamma_2 u}\right)^{\frac{\lambda_2}{2}}. \end{aligned}$$

Therefore the stated result follows. For the last equality in the above expression we have used the integral representation of modified Bessel functions given in Table 4.1. □

Notice that the generality of the GIG distribution translates into a general form for the associated transition probabilities. ○

Other stationary models might be available by suitable constructing bivariate distributions with the same compatible conditional and marginal distributions. However, the drawback of this approach is that the resulting stationary distribution might not fit

the required specifications. In most cases the resulting invariant distribution does not even belong to a well-known parametric family, as in the case of the above example. To some extent, the approach taken by Jørgensen and Song [50] can be seen as a suitable adaptation of this idea, such that underlying invariant distributions meet some specific families.

4.3 STATIONARY ARCH(1) MODELS

In Pitt and Walker [79], a stationary ARCH(1) type model with Student-t distributed errors, was constructed by assuming that the conditional distribution $F_{X|Y}(\cdot \mid y)$ is given by $\mathrm{N}(0, y^{-1})$. This assumption has the following econometric argument: given certain level of volatility "y^{-1}" at time t, the log-returns (X) of a certain generic financial asset at time t are normal distributed. Different models may arise depending on the specification of the marginal distribution for Y. In [79] the marginal for Y was assumed to be $\mathrm{Ga}\left(\nu/2, \nu\beta^2/2\right)$ defining a joint distribution with density

$$f_{X,Y}(x,y) = \mathrm{N}(x; 0, y^{-1})\,\mathrm{Ga}\left(y; \nu/2, \nu\beta^2/2\right) \tag{4.29}$$

and therefore

$$f_{Y|X}(y \mid x) = \mathrm{Ga}\left(y; (\nu+1)/2, (x^2+\nu\beta^2)/2\right).$$

In this case the marginal distribution for X is obtained by integrating out "y" in (4.29), from which we have

$$Q_X = \mathrm{St}(0, \beta^2, \nu)$$

and therefore we can associate a process $\mathrm{X} = \{X_t\}_{t=1}^{\infty}$ with Q_X as its stationary distribution[2]. Hence, a well-defined Markov transition density in discrete time can be constructed as follows:

$$\begin{aligned} p(x_t, x_{t+1}) &= \int_{\mathbb{R}_+} \mathrm{N}(x_{t+1}; 0, y^{-1})\,\mathrm{Ga}\left(y; (\nu+1)/2, (x_t^2+\nu\beta^2)/2\right)\, dy \\ &= \int_{\mathbb{R}_+} \left(\frac{y}{2\pi}\right)^{\frac{1}{2}} \exp\left\{-\frac{x_{t+1}^2 y}{2}\right\} y^{\frac{\nu-1}{2}} \exp\left\{-\frac{y}{2}(\nu\beta^2 + x_t^2)\right\} \frac{\left(\frac{\nu\beta^2+x_t^2}{2}\right)^{\frac{\nu+1}{2}}}{\Gamma(\frac{\nu+1}{2})}\, dy \end{aligned}$$

[2] $\mathrm{St}(\mu, \sigma^2, \nu)$ denotes the non-central Student-t distribution with location parameter μ, dispersion σ and ν degrees of freedom.

$$= \frac{\Gamma\left(\frac{\nu}{2}+1\right)}{\sqrt{2\pi}\Gamma(\frac{\nu+1}{2})} \frac{\left(\frac{\nu\beta^2+x_t^2}{2}\right)^{\frac{\nu+1}{2}}}{\left[\frac{x_{t+1}^2+x_t^2+\nu\beta^2}{2}\right]^{\frac{\nu}{2}+1}} \int_{\mathbb{R}_+} \mathrm{Ga}\left(y; \frac{\nu}{2}+1, \frac{x_{t+1}^2+x_t^2+\nu\beta^2}{2}\right) dy$$

$$= \mathrm{St}\left(x_{t+1}; 0, \frac{{x_t}^2+\nu\beta^2}{\nu+1}, \nu+1\right), \tag{4.30}$$

where $x_t, x_{t+1} \in \mathbb{R}$ and $p(x_t, x_{t+1})$ denotes the Lebesgue density corresponding to $\mathrm{P}(X_{t+1} \in \cdot \mid X_t = x_t)$.

With the transition density (4.30) it is possible to construct a version of the following ARCH(1) time series model

$$X_{t+1} = \sqrt{\frac{X_t^2+\nu\beta^2}{\nu+1}}\varepsilon_{t+1}, \quad \varepsilon_{t+1} \sim \mathrm{St}(0,1,\nu+1). \tag{4.31}$$

Model (4.31) can be identified as an AR(1) model in X_t^2, that is, an ARCH(1) model with errors Student-t distributed. By construction, the existence of the required stationary distribution follows. If $X_t \sim \mathrm{St}(0,\beta^2,\nu)$ and X_{t+1} given X_t follows the transition (4.30), then $X_{t+1} \sim \mathrm{St}(0,\beta^2,\nu)$. See Pitt and Walker [79].

Clearly, the existence of a model with the required stationary distribution does not necessarily imply its uniqueness. See Example 3.2. However, in this case, the advantage of this model is its identifiability as an ARCH process and therefore the availability of the latent structure decomposition. This, will ease some estimation issues as we will see in Section 4.10.

Model (4.31), is the stationary version of a model first studied by Bollerslev [17], given by

$$X_{t+1} = \sqrt{a+bX_t^2}\varepsilon_{t+1}, \quad \varepsilon_{t+1} \sim \mathrm{St}(0,1,m).$$

Model (4.31) arises when $a = \nu\beta^2$, $b = 1/(1+\nu)$ and constraining $bm = 1$. Here, a is regarded as the average level of volatility and b as the dynamic variability in the variance. Notice that the construction presented in [79] ensures the stationarity of the model in contrast to the model presented in [17].

Clearly, the above construction inherited the properties of a sequence generated via the Gibbs sampler method, such as the hypothesized reversibility. Here, it is worth

noting that X can be thought as the observed process and Y the latent (unobservable) process. As we will see in Section 4.11, X will be associated with log-returns over time, and Y as the underlying volatility.

The conditional moments for model (4.31) can be easily obtained via the characteristic function of (4.30). First, let us notice that if $X \sim \text{Iga}(\alpha, \beta)$, then the Laplace transform corresponding to X, is given by

$$\mathcal{L}_X(\theta) = \int_0^\infty e^{-\theta x - \frac{\beta}{x}} \frac{\beta^\alpha}{\Gamma(\alpha)} x^{-(\alpha+1)}\, dx = \frac{-2[\sqrt{\theta\beta}]^\alpha}{\Gamma(\alpha)} K_\alpha(2\sqrt{\theta\beta}), \tag{4.32}$$

where K_λ denotes the modified Bessel function of the third kind with index λ. For the above integral we have used the integral representation of Bessel functions presented in Table 4.1. If we notice that

$$\text{St}(y; \mu, \sigma^2, \nu) = \int_{-\infty}^\infty \text{N}(y; \mu, \sigma^2 x)\, \text{Iga}\left(x; \frac{\nu}{2}, \frac{\nu}{2}\right) dx \tag{4.33}$$

then we can find the characteristic function corresponding to $Y \sim \text{St}(\mu, \sigma^2, \nu)$ as follows

$$\begin{aligned} \mathcal{C}_Y(\xi) &= \int_0^\infty \exp\{i\xi\mu - \xi^2\sigma^2 x/2\} \text{Iga}\left(x; \frac{\nu}{2}, \frac{\nu}{2}\right) dx \\ &= e^{i\xi\mu} \mathcal{L}_X(\xi^2\sigma^2/2) \\ &= -\frac{2e^{i\xi\mu}}{\Gamma(\nu/2)} \left(\frac{\sqrt{\xi^2\sigma^2\nu}}{2}\right)^{\nu/2} K_{\nu/2}\left(\sqrt{\xi^2\sigma^2\nu}\right). \end{aligned} \tag{4.34}$$

Hence, the moments for Y, when available[3], can be found by $\mathbb{E}[Y^k] = \mathcal{C}_Y^{(k)}(0)/i^k$. For example,

$$\begin{aligned} \text{Cov}(X_{t+1}^2, X_t^2) &= \mathbb{E}\left\{X_t^2 \mathbb{E}[X_{t+1}^2 \mid X_t]\right\} - \mathbb{E}[X_{t+1}^2]\mathbb{E}[X_t^2] \\ &= \frac{\mathbb{E}\left[X_t^2(X_t^2 + \nu\beta^2)\right]}{\nu+1} - \mathbb{E}[X_{t+1}^2]\mathbb{E}[X_t^2] \\ &= \frac{\mathbb{E}[X_t^4]}{\nu-1} + \frac{\nu\beta^2\mathbb{E}[X_t^2]}{\nu-1} - \mathbb{E}[X_t^2]^2 \\ &= 2\frac{\nu^2\beta^4}{(\nu-4)(\nu-2)^2}, \end{aligned}$$

where the third equality is due to the stationarity of $\{X_t\}_{t=1}^\infty$. In the same way we see

[3]Student-t distribution embrace the Cauchy distribution, for which moments does not exist.

that

$$\begin{aligned} \mathrm{Var}(X_t^2) &= \mathbb{E}[X_t^4] - \mathbb{E}[X_t^2]^2 \\ &= 2\frac{\nu^2\beta^4(\nu-1)}{(\nu-4)(\nu-2)^2}, \end{aligned}$$

implying that the autocorrelation function is given by

$$\rho(h) = \mathrm{Corr}(X_t^2, X_{t+h}^2) = (\nu-1)^{-h}, \tag{4.35}$$

provided that $\nu > 4$, so that the fourth moment exists.

Also notice that $\mathrm{Corr}(X_t, X_{t+h}) = 0$ for any h, this agrees with the common assumption that log-returns are uncorrelated whereas their squares exhibit serial correlation. On the other hand, this same observation emphasizes the fact that the reversibility property is a higher moment (higher than 2) dependence property as mentioned in Section 3.3.

4.4 GENERALIZED-HYPERBOLIC ARCH(1) STATIONARY MODEL

The construction in Section 4.3 can be generalized by either assuming a more general family for $F_{Y|X}$ or by assuming a more general conjugate family to Q_Y.

The Student-t distribution results from a mixture of normal and inverse gamma distributions. Inverse gamma distributions can be seen as particular case of generalized inverse Gaussian (GIG) distributions, with parameters $\gamma = 0$ and $\lambda < 0$, see equation (4.16). As we have seen in equation (4.14) a normal variance-mean mixture with a GIG distribution as a mixing distribution results in generalized hyperbolic (GH) distributions. Furthermore, the Student-t distribution can be seen as a particular case of GH distribution. For this reason, it is intuitive to use the GIG distribution to generalize the Inverse gamma distribution as a model for Q_Y. Using a similar construction to the one presented for the Student-t case we get a generalized hyperbolic stationary ARCH(1) process. This generalization allows better modelling for financial log-returns. ARCH models with GH errors have been previously introduced in Barndorff-Nielsen [9], however the difference in our approach lies in that the constructed model is stationary.

Let us generalize the assumption for $F_{X|Y}$ and Q_Y by assuming

$$F_{X|Y}(\cdot \mid y) = \mathrm{N}(\mu + \beta y, y) \quad \text{and} \quad Q_Y(\cdot) = \mathrm{GIG}(\lambda, \delta^2, \alpha^2 - \beta^2), \tag{4.36}$$

where $\mu, \beta, \lambda \in \mathbb{R}$ and $0 \leq |\beta| < \alpha$. Again the process X models the log-returns of a certain generic financial asset at time t given certain level of volatility "y". Consequently, we have

$$\begin{aligned} f_{Y|X}(y \mid x) &\propto y^{\lambda - \frac{3}{2}} \exp\left\{ -\frac{(x - (\mu - \beta y))^2}{2y} - \frac{(\delta^2 y^{-1} + (\alpha^2 - \beta^2) y)}{2} \right\} \\ &\propto y^{\lambda - \frac{3}{2}} \exp\left\{ -\frac{1}{2} \left[((x-\mu)^2 + \delta^2) y^{-1} + \alpha^2 y \right] \right\}. \end{aligned}$$

Normalizing the above quantity, we obtain

$$f_{Y|X}(y \mid x) = \mathrm{GIG}\left(y; \lambda - \frac{1}{2}, (x-\mu)^2 + \delta^2, \alpha^2 \right).$$

With the given conditional probabilities, the one-step transition probability is given by

$$\begin{aligned} p(x_t, x_{t+1}) &= \int_{\mathbb{R}_+} \frac{\left(\frac{\alpha^2}{(x_t - \mu)^2 + \delta^2} \right)^{\frac{\lambda - 1/2}{2}} y^{\lambda - 2} \exp\{\beta(x_{t+1} - \mu)\}}{2\sqrt{2\pi}\, K_{\lambda - \frac{1}{2}}\left(\alpha \sqrt{(x_t - \mu)^2 + \delta^2} \right)} \\ &\times \exp\left\{ -\frac{\left[(x_{t+1} - \mu)^2 + (x_t - \mu)^2 + \delta^2 \right]}{2y} - \frac{(\beta^2 + \alpha^2) y}{2} \right\} dy \\ &= a\left(\lambda - \frac{1}{2}, \sqrt{\beta^2 + \alpha^2}, \beta, \sqrt{(x_t - \mu)^2 + \delta^2} \right) \exp\{\beta(x_{t+1} - \mu)\} \\ &\times \left\{ \sqrt{(x_{t+1} - \mu)^2 + (x_t - \mu)^2 + \delta^2} \right\}^{\lambda - 1} \\ &\times K_{\lambda - 1}\left(\sqrt{(\beta^2 + \alpha^2)((x_{t+1} - \mu)^2 + (x_t - \mu)^2 + \delta^2)} \right) \\ &= \mathrm{GH}\left(x_{t+1}; \lambda - \frac{1}{2}, \sqrt{\beta^2 + \alpha^2}, \beta, \sqrt{(x_t - \mu)^2 + \delta^2}, \mu \right). \end{aligned} \tag{4.37}$$

Proposition 4.2. If $X_t \sim \mathrm{GH}(\lambda, \alpha, \beta, \delta, \mu)$ and the transition density $p(x_t, x_{t+1})$ is given by (4.37), then marginally $X_{t+1} \sim \mathrm{GH}(\lambda, \alpha, \beta, \delta, \mu)$.

Proof. By construction it would suffice to get the marginal for X using the joint distribution formed by $f_{X|Y}(x, y) q_Y(y)$. However, we can equally proceed as follows:

$$q_X(x_{t+1}) = \int_{\mathbb{R}} p(x_t, x_{t+1})\, q_X(x_t)\, dx_t$$

$$
\begin{aligned}
&= \int_{\mathbb{R}} \mathrm{GH}\left(x_{t+1}; \lambda - \frac{1}{2}, \sqrt{\beta^2+\alpha^2}, \beta, \sqrt{(x_t-\mu)^2+\delta^2}, \mu\right) \mathrm{GH}\left(x_t; \lambda, \alpha, \beta, \delta, \mu\right) dx_t \\
&= \frac{K_{\lambda-\frac{1}{2}}\left(\alpha\sqrt{(x_{t+1}-\mu)^2+\delta^2}\right)\left(\sqrt{(x_{t+1}-\mu)^2+\delta^2}\right)^{\lambda-\frac{1}{2}} (\alpha^2-\beta^2)^{\lambda/2}}{\sqrt{2\pi}\,\alpha^{\lambda-\frac{1}{2}}\,\delta^{\lambda}\,K_{\lambda}\left(\delta\sqrt{\alpha^2-\beta^2}\right)\exp\left\{-\beta(x_{t+1}-\mu)\right\}} \\
&\times \int_{\mathbb{R}} \frac{\left[(x_{t+1}-\mu)^2+(x_t-\mu)^2+\delta^2\right]^{\frac{\lambda-1}{2}} (\alpha^2)^{\frac{\lambda-1/2}{2}} \exp\{\beta(x_t-\mu)\}}{\left(\sqrt{\beta^2+\alpha^2}\right)^{\lambda-1}\sqrt{2\pi}\left(\sqrt{(x_{t+1}-\mu)^2+\delta^2}\right)^{\lambda-\frac{1}{2}}} \\
&\times \frac{K_{\lambda-1}\left(\sqrt{(\beta^2+\alpha^2)\left[(x_{t+1}-\mu)^2+(x_t-\mu)^2+\delta^2\right]}\right)}{K_{\lambda-\frac{1}{2}}\left(\alpha\sqrt{(x_{t+1}-\mu)^2+\delta^2}\right)} dx_t \qquad (4.38) \\
&= \mathrm{GH}\left(x_{t+1}; \lambda, \alpha, \beta, \delta, \mu\right) \int_{\mathbb{R}} \mathrm{GH}\left(x_t; \lambda - \frac{1}{2}, \sqrt{\beta^2+\alpha^2}, \beta, \sqrt{(x_{t+1}-\mu)^2+\delta^2}, \mu\right) dx_t \\
&= \mathrm{GH}\left(x_{t+1}; \lambda, \alpha, \beta, \delta, \mu\right). \qquad (4.39)
\end{aligned}
$$

Comparing the second equality with the fourth one, in the above expression, the reversibility property is verified. □

Proposition 4.3. The correlation for the stationary generalized hyperbolic ARCH(1) model is given by

$$
\mathrm{Corr}(X_{t+1}, X_t) = \frac{\mathrm{Cov}(X_{t+1}, X_t) K_{\lambda}(\omega)}{\eta K_{\lambda+1}(\omega) + \mathrm{Cov}(X_{t+1}, X_t) K_{\lambda}(\omega)},
$$

where

$$
\mathrm{Cov}(X_{t+1}, X_t) = \beta^2\eta^2 \left\{ \frac{K_{\lambda+2}(\omega)}{K_{\lambda}(\omega)} - \left(\frac{K_{\lambda+1}(\omega)}{K_{\lambda}(\omega)}\right)^2 \right\}
$$

and $\omega = \delta\sqrt{\alpha^2-\beta^2}$ and $\eta = \delta/\sqrt{\alpha^2-\beta^2}$.

Proof. The stationarity of the process $\{X_t\}_{t=1}^{\infty}$ implies

$$
\begin{aligned}
\mathrm{Cov}(X_{t+1}, X_t) &= \mathbb{E}[X_{t+1}X_t] - \mathbb{E}[X_t]^2 \\
&= \mathbb{E}\{X_t \mathbb{E}[X_{t+1} \mid X_t]\} - \mathbb{E}[X_t]^2. \qquad (4.40)
\end{aligned}
$$

For notation simplicity we let $r_x = \sqrt{(x-\mu)^2+\delta^2}$, $\omega_x = \alpha r_x$ and $\eta_x = r_x/\alpha$ be a re-parametrization corresponding to the transition probability (4.37).

First let us work out the cross moment in expression (4.40)

$$\begin{aligned}\mathbb{E}\{X_t\mathbb{E}[X_{t+1} \mid X_t]\} &= \mathbb{E}\left\{X_t\left(\mu+\beta\eta_{X_t}\frac{K_{\lambda+\frac{1}{2}}(\omega_{X_t})}{K_{\lambda-\frac{1}{2}}(\omega_{X_t})}\right)\right\}\\ &= \mu\,\mathbb{E}[X_t]+\beta\,\mathbb{E}\left\{X_t\eta_{X_t}\frac{K_{\lambda+\frac{1}{2}}(\omega_{X_t})}{K_{\lambda-\frac{1}{2}}(\omega_{X_t})}\right\}.\end{aligned}$$

Where,

$$\begin{aligned}\mathbb{E}\left[X_t\eta_{X_t}\frac{K_{\lambda+\frac{1}{2}}(\omega_{X_t})}{K_{\lambda-\frac{1}{2}}(\omega_{X_t})}\right] &= \frac{a(\lambda,\alpha,\beta,\delta)}{\alpha a(\lambda+1,\alpha,\beta,\delta)}\\ &\times \int_{\mathbb{R}} x_t \underbrace{a(\lambda+1,\alpha,\beta,\delta)\,r_{x_t}^{\lambda+\frac{1}{2}}K_{\lambda+\frac{1}{2}}(\omega_{x_t})\,e^{\beta(x_t-\mu)}}_{\mathrm{GH}(\lambda+1,\alpha,\beta,\delta,\mu)}dx_t\\ &= \eta\frac{K_{\lambda+1}(\omega)}{K_\lambda(\omega)}\,\mathbb{E}_{\lambda+1}[X_t] \qquad (4.41)\\ &= \eta\frac{K_{\lambda+1}(\omega)}{K_\lambda(\omega)}\left\{\mu+\beta\eta\frac{K_{\lambda+2}(\omega)}{K_{\lambda+1}(\omega)}\right\}\\ &= \mu\eta\frac{K_{\lambda+1}(\omega)}{K_\lambda(\omega)}+\beta\eta^2\frac{K_{\lambda+2}(\omega)}{K_\lambda(\omega)}.\end{aligned}$$

In equation (4.41) the expectation $\mathbb{E}_{\lambda+1}[\cdot]$ is taken with respect to $\mathrm{GH}(\lambda+1,\alpha,\beta,\delta,\mu)$. Here is worth noticing that in general for $n=0,1,2,\ldots$ and any function f we have

$$\mathbb{E}_\lambda\left[f(X_t)\eta_{X_t}^{n+1}\frac{K_{\lambda+n+\frac{1}{2}}(\omega_{X_t})}{K_{\lambda-\frac{1}{2}}(\omega_{X_t})}\right] = \eta^{n+1}\frac{K_{\lambda+n+1}(\omega)}{K_\lambda(\omega)}\,\mathbb{E}_{\lambda+n+1}[f(X_t)].$$

Therefore, we can write

$$\mathbb{E}[X_{t+1}X_t] = \mu\,\mathbb{E}[X_t]+\beta\eta\,\frac{K_{\lambda+1}(\omega)}{K_\lambda(\omega)}\,\mathbb{E}_{\lambda+1}[X_t]. \qquad (4.42)$$

Substituting the marginal expectation $\mathbb{E}[X_t]$ (equation 4.21) into equation (4.40) we get the stated result for the covariance $\mathrm{Cov}(X_{t+1},X_t)$. Finally, using the second marginal moment (4.22) we get

$$\begin{aligned}\mathrm{Var}(X_t) &= \mathbb{E}[X_t^2]-\mathbb{E}[X_t]^2\\ &= \frac{K_{\lambda+1}(\omega)}{K_\lambda(\omega)}\eta+\beta^2\eta^2\left\{\frac{K_{\lambda+2}(\omega)}{K_\lambda(\omega)}-\left(\frac{K_{\lambda+1}(\omega)}{K_\lambda(\omega)}\right)^2\right\}\\ &= \frac{K_{\lambda+1}(\omega)}{K_\lambda(\omega)}\eta+\mathrm{Cov}(X_{t+1},X_t). \qquad (4.43)\end{aligned}$$

The underlying stationarity implies

$$\text{Corr}(X_{t+1}, X_t) = \frac{\text{Cov}(X_{t+1}, X_t)}{\text{Var}(X_t)}. \tag{4.44}$$

A direct substitution of equation (4.43) into (4.44) leads to the result. □

For the symmetric case, that is when $\beta = 0$, the correlation is clearly 0. This agrees with the empirical observation that financial log-returns are zero correlated. An important property to underline here is that the general form of our GH-ARCH(1) model allows for asymmetric invariant distributions, a property that is not common among most of the reversible models in the literature. When we allow for asymmetry the correlation structure gets more complicated and even the non-squared process shows correlation.

Another quantity of interest is the correlation in the squares, that is, $\text{Corr}(X_{t+1}^2, X_t^2)$. In order to compute this quantity we can proceed as in the proof for Proposition 4.3. That is, we can compute

$$\text{Corr}(X_{t+1}^2, X_t^2) = \frac{\mathbb{E}[X_{t+1}^2 X_t^2] - \mathbb{E}[X_t^2]^2}{\mathbb{E}[X_t^4] - \mathbb{E}[X_t^2]^2}, \tag{4.45}$$

where the second and fourth moments are given by equations (4.22) and (4.24) respectively. The underlying cross moment can be found as follows

$$\begin{aligned} &\mathbb{E}[X_{t+1}^2 X_t^2] = \mathbb{E}\left\{X_t^2 \mathbb{E}[X_{t+1}^2 \mid X_t]\right\} \qquad (4.46)\\ = \;& \mathbb{E}\left\{X_t^2 \left(\mu^2 + \frac{K_{\lambda+\frac{1}{2}}(\omega_{X_t})}{K_{\lambda-\frac{1}{2}}(\omega_{X_t})}\eta_{X_t}(2\mu\beta+1) + \frac{K_{\lambda+\frac{3}{2}}(\omega_{X_t})}{K_{\lambda-\frac{1}{2}}(\omega_{X_t})}\eta_{X_t}^2\beta^2\right)\right\}\\ = \;& \mu^2\,\mathbb{E}_\lambda[X_t^2] + (2\mu\beta+1)\eta\frac{K_{\lambda+1}(\omega)}{K_\lambda(\omega)}\,\mathbb{E}_{\lambda+1}[X_t^2] + \beta^2\eta^2\frac{K_{\lambda+2}(\omega)}{K_\lambda(\omega)}\,\mathbb{E}_{\lambda+2}[X_t^2]. \end{aligned}$$

Note the peculiar relation of the cross moment of the squares (4.46) with the marginal second moment (4.22), as well as the relation of the simple cross moment (4.42) with the marginal expectation (4.24). This, suggests a general formula, similar to the one presented in Theorem 4.1, characterizing all the moments for a generalized hyperbolic distribution.

Apparently, the general expression for the correlation (4.45) does not have a short form, therefore we have not displayed it here. However, we turn to the simpler, though not

less important, symmetric case, for which the correlation is given by

$$\mathrm{Corr}_{GH(\lambda,\alpha,0,\delta,\mu)}\left(X_{t+1}^2, X_t^2\right) = \frac{\delta\left\{K_{\lambda+2}(\alpha\delta)\,K_\lambda(\alpha\delta) - (K_{\lambda+1}(\alpha\delta))^2\right\}}{4\mu^2\alpha\,K_{\lambda+1}(\alpha\delta)\,K_\lambda(\alpha\delta) + 3\delta\,K_{\lambda+2}(\alpha\delta)\,K_\lambda(\alpha\delta) - (K_{\lambda+1}(\alpha\delta))^2}.$$

In particular, for $\lambda = -1/2$

$$\mathrm{Corr}_{NIG(\alpha,0,\delta,\mu)}\left(X_{t+1}^2, X_t^2\right) = \frac{1}{4\mu^2\alpha^2 + 2\delta\alpha + 3}.$$

Figure 4.3 gives an example of how the existence of a skewness parameter (β) may allow for bigger correlations. The graph is presented for different values of λ.

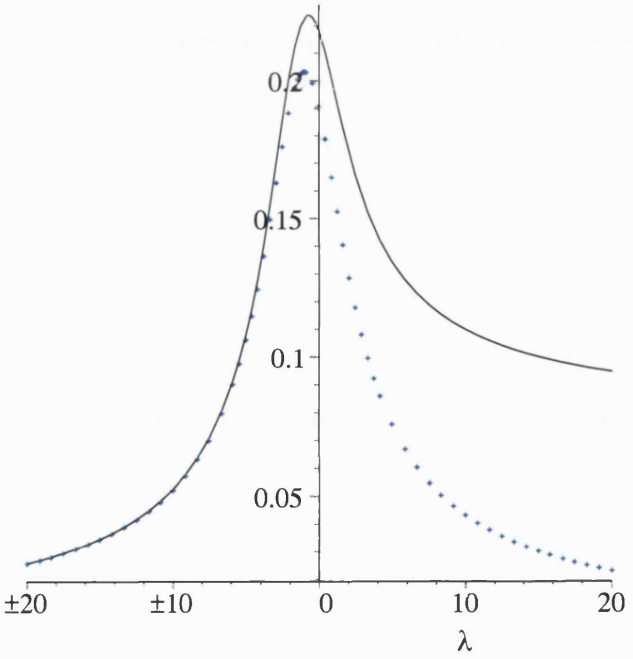


Figure 4.3: One-lag correlation for the squares in a stationary GH-ARCH(1) model. The solid line corresponds to a model with parameters $\alpha = 1, \beta = -0.2, \delta = 1, \mu = 0$ and the other line to a model with parameters $\alpha = 1, \beta = 0, \delta = 1, \mu = 0$.

4.5 P-LAGGED TIME SERIES

So far, we have only constructed time series models with one lag-dependence. Now we will describe a straightforward generalization to this construction. Instead of considering the two-dimensional distributions as in (3.12), let us consider the $(p+1)$-dimensional distributions given by

$$p(X^{(t,p)}) = q(X_t)\prod_{i=1}^{p} p(X_{t+i} \mid X^{(t,i-1)}), \tag{4.47}$$

where $X^{(t,i)} := (X_t, \ldots, X_{t+i})$ for any $t \in \mathbb{N}$. As before, in order to keep the strictly stationarity of the sequence $\{X_t\}_{t=1}^{\infty}$ with *fdds* given by equation (4.47), we need to impose further conditions on the updating mechanism $p(X_{t+i} \mid X^{(t,i-1)})$. That is, we need to ensure the symmetry of the p-dimensional distributions given the invariant distribution Q_X. This can be equally done via a Gibbs sampler type updating. Assume that this updating mechanism is given by

$$\begin{aligned}
&(i) \quad \{Y_{t+i} \mid X^{(t,i-1)} = x^{(t,i-1)}, Y^{(t,i-1)} = y^{(t,i-1)}\} \\
&\quad \stackrel{\mathrm{d}}{=} \{Y_{t+i} \mid X^{(t,i-1)} = x^{(t,i-1)}\} \qquad\qquad \sim F_{Y|X^i}(\cdot \mid x^{(t,i-1)}) \\
&(ii) \quad \{X_{t+i} \mid Y_{t+i} = y, X^{(t,i-1)} = x^{(t,i-1)}, Y^{(t,i-1)} = y^{(t,i-1)}\} \\
&\quad \stackrel{\mathrm{d}}{=} \{X_{t+i} \mid Y_{t+i} = y_{t+i}\} \qquad\qquad \sim F_{X|Y}(\cdot \mid y_{t+i}),
\end{aligned} \tag{4.48}$$

where $i = 1, \ldots, p$, X^i denotes an i-dimensional random vector and $x^{(t,i-1)}$ an i-dimensional vector denoting the time-space values corresponding to $X^{(t,i-1)}$. If $p = 1$, we get the same case described in Example 3.2 or in Section 4.3. Due to the underlying conditional independence and under the knowledge of Q_Y, the density for $F_{Y|X^i}$ can be found as follows

$$f_{Y|X^i}\left(y_{t+i} \mid x^{(t,i-1)}\right) \propto q_Y(y_{t+i}) \prod_{j=1}^{i} f_{X|Y}(x_{t+j-1} \mid y_{t+j}). \tag{4.49}$$

Therefore, assuming that marginally $X_t \sim Q_X$ and using the conditional independence structure assumed in (4.48), we only need to specify the forms of $F_{Y|X^i}$ for $i = 1, \ldots, p$ in order to have a strictly stationary process with invariant measure Q_X. As before, the

specification of the functional forms for $F_{Y|X^i}$ is quite open. The associated one-step transition density for this model is given by

$$p(x_{t+i} \mid x^{(t,i-1)}) = \int_E f_{X|Y}(x_{t+i} \mid y)\, f_{Y|X^i}(y \mid x^{(t,i-1)})\, \lambda_Y(dy). \tag{4.50}$$

Following Pitt and Walker [79], a similar form for the conditional distributions can be introduced. The next section generalizes their construction.

4.6 STUDENT-T ARCH(p) STATIONARY MODEL

For the sake of illustration we start with the Student-t ARCH case. We consider the same arguments used in Section 4.3 to construct the p-lagged updating scheme (4.48). We can construct the conditional distribution $F_{Y|X^i}$ following expression (4.49), first let us notice that

$$f_{X^p|Y}(x^p \mid y) = \prod_{j=1}^{p} f_{X_j|Y}(x_j \mid y) = \mathrm{N}_p(x^p; 0, y^{-1}\mathrm{I}), \tag{4.51}$$

where $\mathrm{N}_p(\cdot,\cdot)$ represents the p-dimensional normal distribution and I the identity matrix. Notice that in general $F_{X^i|Y}$ may not enjoy the simplicity implied by the normality assumptions in (4.51).

Again assuming $\mathrm{Ga}\left(\nu/2, \nu\beta^2/2\right)$ as the marginal distribution of Y, we can write the updating mechanism in (4.48) as follows

$$\begin{aligned} &(i) \qquad \{Y_{t+i} \mid X^{(t,i-1)} = x^{(t,i-1)}\} \quad \sim \mathrm{Ga}\left(\tfrac{\nu+i}{2}, \tfrac{\nu\beta^2+\sum_{j=1}^{i} x_{t+i-j}^2}{2}\right) \\ &(ii) \qquad \{X_{t+i} \mid Y_{t+i} = y_{t+i}\} \qquad\quad \sim \mathrm{N}(0, y_{t+i}^{-1}), \end{aligned} \tag{4.52}$$

for $i = 1, \ldots, p$. Integrating out y_{t+i}, which might represent certain level of volatility at time $t+i$, the one-step transition probability (4.50) simplifies as follows

$$p(x_{t+i} \mid x^{(t,i-1)}) = \mathrm{St}\left(x_{t+i}; 0, \frac{\nu\beta^2 + \sum_{j=1}^{i} x_{t+i-j}^2}{\nu+i}, \nu+i\right). \tag{4.53}$$

In order to get the above equality we only need to multiply by the constant that makes

$$\mathrm{Ga}\left(\frac{\nu+i+1}{2}, \frac{\nu\beta^2+\sum_{j=0}^{i} x_{t+i-j}^2}{2}\right)$$

integrate out one.

Notice that here the updating mechanism in (4.52) is based on the joint distribution

$$f_{X^p,Y}(x^p, y) = \mathrm{Ga}\left(y; \nu/2, \nu\beta^2/2\right) \mathrm{N}_p(x^p; 0, y^{-1}I). \tag{4.54}$$

Proposition 4.4. If $X_t \sim \mathrm{St}(0, \beta^2, \nu)$ and $X_{t+i} \mid X^{(t,i-1)}$ follows (4.53) for $i = 1, \ldots, p$, then marginally $X_{t+p} \sim \mathrm{St}(0, \beta^2, \nu)$.

Proof. The simplest argument is given by construction. Notice that the marginal distribution for X in (4.54) is given by the required invariant distribution $\mathrm{St}(0, \beta^2, \nu)$. However, in this case we verify this fact by using the reversibility property of the transition probability and by a simple application of Fubini's theorem.

$$\begin{aligned}
q_X(x_{t+p}) &= \int_{\mathbb{R}^p} q_X(x_t) \prod_{i=0}^{p-1} p(x_{t+p-i} \mid x_{t+p-i-1})\, dx_{t+p-1} \cdots dx_t \\
&= \int_{\mathbb{R}^p} q_X(x_{t+p}) \prod_{i=0}^{p-1} p(x_{t+p-i-1} \mid x_{t+p-i})\, dx_t \cdots dx_{t+p-1} \\
&= \mathrm{St}(x_{t+p}; 0, \beta^2, \nu) \int_{\mathbb{R}} \cdots \int_{\mathbb{R}} \mathrm{St}\left(x_{t+p-1}; 0, \frac{\nu\beta^2 + x_{t+p}^2}{\nu+1}, \nu+1\right) \\
&\times \left[\prod_{i=1}^{p-2} \mathrm{St}\left(x_{t+i}; 0, \frac{\nu\beta^2 + \sum_{j=0}^{p-i-1} x_{t+p-j}^2}{\nu+p-i}, \nu+p-i\right)\right] \\
&\times \mathrm{St}\left(x_t; 0, \frac{\nu\beta^2 + \sum_{j=0}^{p-1} x_{t+p-j}^2}{\nu+p}, \nu+p\right) dx_t dx_{t+1} \cdots dx_{t+p-1} \\
&= \mathrm{St}(x_{t+p}; 0, \beta^2, \nu).
\end{aligned} \tag{4.55}$$

□

Proposition 4.4 generalizes the model (4.31) introduced by Pitt and Walker [79]. In

this case the model can be represented as follows

$$X_{t+p} = \sqrt{\frac{\nu\beta^2 + \sum_{j=1}^{p} X_{t+p-j}^2}{\nu + p}}\,\varepsilon_{t+p}, \quad \varepsilon_{t+p} \sim \mathrm{St}(0, 1, \nu + p). \tag{4.56}$$

Here, the updated mechanism needed in order to construct the *fdds* $p(X_t, X_{t+1}, \ldots, X_{t+p})$, using a simplified notation, can be written as

$$\begin{aligned}
X_{t+1} &\mid X_t \\
X_{t+2} &\mid X_t, X_{t+1} \\
&\vdots \\
X_{t+j} &\mid X_t, X_{t+1}, \ldots, X_{t+j-1} \\
&\vdots \\
X_{t+p} &\mid X_t, X_{t+1}, \ldots, X_{t+i}, \ldots, X_{t+p-1},
\end{aligned}$$

where each conditional is assumed to follow equation (4.53) and $X_t \sim Q_X$.

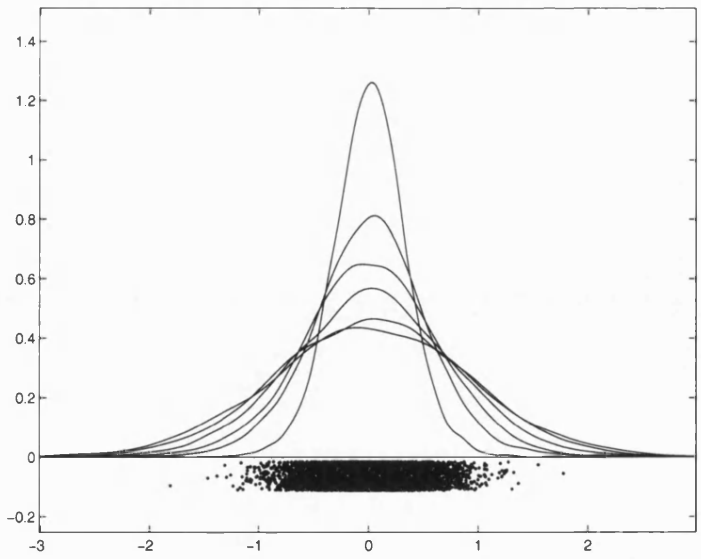


Figure 4.4: Densities from 10000 simulated St-ARCH(10) observations, $\nu = 1, 3, 5, 10, 20, 40$ and $\beta = 1$. The densities were smoothed using a nonparametric estimate with gaussian kernel. The more leptokurtic the density is the less degrees of freedom.

It is clear, from model (4.56), that for $p > 1$, X does not satisfy the Markov property, because of the dependence on the last p observations. However, a new p-order Markov

process can be associated. In Figure 4.4 the density estimates for some simulations of an St-ARCH(10) model are presented.

One of the disadvantages of the above construction is that we cannot have a different value for each of the past lagged values, as in the case of Engle's [32] ARCH process. Therefore, the effects of the past values considered in the model are assumed to have the same weight. In order to clarify this let us consider the particular case where $p = 2$. In this case, using (4.53), we have

$$p(x_{t+2} \mid x_{t+1}, x_t) = \mathrm{St}\left(x_{t+2}; 0, \frac{x_{t+1}^2 + x_t^2 + \nu\beta^2}{\nu + 2}, \nu + 2\right).$$

On the other hand, if we allow the parameter β to change in time, we could have for example the following scheme: let us assume that

$$f_{X|Y}(x_{t+i-1} \mid y_{t+i}) = \mathrm{N}\left(x_{t+i-1}; 0, \beta_i^2/y_{t+i}\right)$$

and

$$q_Y(y_{t+i}) = \mathrm{Ga}\left(y_{t+i}; \frac{\nu}{2}, \frac{\nu}{2}\right),$$

where $i = 1, 2$, then from (4.49)

$$f_{Y|X^{(2)}}(y_{t+2} \mid x_{t+1}, x_t) = \mathrm{Ga}\left(y_{t+i}; \frac{\nu}{2} + 1, \frac{1}{2}\left[\frac{x_{t+1}^2}{\beta_1^2} + \frac{x_t^2}{\beta_0^2} + \nu\right]\right).$$

Therefore we have the following transition density

$$\begin{aligned} p\left(x_{t+2} \mid x_{t+1}, x_t\right) &= \int f(x_{t+2} \mid y_{t+i})\, f(y_{t+i} \mid x_{t+1}, x_t)\, \lambda_y(dy_{t+i}) \qquad (4.57) \\ &= \mathrm{St}\left(x_{t+2}; 0, \beta_2^2 \left[\frac{\frac{x_{t+1}^2}{\beta_1^2} + \frac{x_t^2}{\beta_0^2} + \nu}{\nu + 2}\right], \nu + 2\right). \end{aligned}$$

In a similar way we can see that

$$p\left(x_{t+1} \mid x_t\right) = \mathrm{St}\left(x_{t+1}; 0, \beta_1^2 \left[\frac{\frac{x_t^2}{\beta_0^2} + \nu}{\nu + 1}\right], \nu + 1\right).$$

In this case, if we assume that $X_t \sim \mathrm{St}(0, \beta_0^2, \nu)$, the corresponding marginal density is

given by

$$\begin{aligned} q_X(x_{t+2}) &= \mathrm{St}(x_{t+2}; 0, \beta_2^2, \nu) \int \mathrm{St}\left(x_{t+1}; 0, \beta_1^2 \left[\frac{\frac{x_{t+2}^2}{\beta_2^2} + \nu}{\nu + 1} \right], \nu + 1 \right) \\ &\times \int \mathrm{St}\left(x_t; 0, \beta_0^2 \left[\frac{\frac{x_{t+2}^2}{\beta_2^2} + \frac{x_{t+1}^2}{\beta_1^2} + \nu}{\nu + 2} \right], \nu + 2 \right) dx_t dx_{t+1} \\ &= \mathrm{St}(x_{t+2}; 0, \beta_2^2, \nu). \end{aligned}$$

As we can see, the stationarity condition does not hold, unless $\beta_0 = \beta_1 = \beta_2$ which leads to the same case discussed before in Proposition 4.4.

4.7 GENERALIZED HYPERBOLIC ARCH(p) STATIONARY MODEL

The use of the GH distribution instead of the Student-t distribution for ARCH models was first suggested by Barndorff-Nielsen [9]. Although he proposed a more general case than the model we are going to present, using a general function $r(X^{(t,i)}, \eta)$ for the lag-dependence observations $X^{(t,i)}$ and some parameters represented by η, this does not always lead to the stationary case. For example, in [9] the case

$$r(X^{(t,i)}, \eta) = \left(\varepsilon + \sum_{j=0}^{i} \rho_j X_{t+j}^2 \right)^{1/2} \tag{4.58}$$

was suggested, leading in general to non-stationary models. Intuitively, as we have seen for the Student-t case, different weights for each lagged values destroys the "strictly" stationarity property.

The objective here is to find a function $r(\cdot, \cdot)$ that leads to a stationary GH-ARCH(p) model, just as in the case of the Student-t ARCH(p). To construct the conditional distribution $F_{Y|X}$ we assume that (4.36) holds, as we did for the GH-ARCH(1). With this specification and using construction (4.49) we notice that

$$f_{X^p|Y}(x^p \mid y) = \mathrm{N}_p\left(x^p; \mathrm{M}_p + y\mathrm{B}_p, y\mathrm{I}\right), \tag{4.59}$$

where $\mathrm{M}_p, \mathrm{B}_p \in \mathbb{R}^p$ and I denotes the identity matrix.

In fact, from the outset, a more general model can be considered

$$f_{X^p|Y}(x^p \mid y) = \mathrm{N}_p\left(x^p; \mathrm{M}_p + y\Delta_p \mathrm{B}_p, y\Delta_p\right), \tag{4.60}$$

where $\Delta_p \in \mathbb{R}^{p\times p}$ is a positive definite matrix with $|\Delta_p| = 1$. The specification of Q_Y is given by

$$Q_Y(\cdot) = \mathrm{GIG}(\lambda, \delta^2, \alpha^2 - \mathrm{B}_p^T \Delta_p \mathrm{B}_p), \tag{4.61}$$

where $\lambda \in \mathbb{R}$, $\delta > 0$, $\Delta_p \in \mathbb{R}^{p\times p}$ and $\alpha^2 > B_p^T \Delta_p B_p$. With the conditional (4.60) and equation (4.61) we can compute the conditional density (4.49), getting the following:

$$\begin{aligned} & f_{Y|X^p}(y \mid x^p) \propto y^{\lambda-\frac{p}{2}-1} \\ \times \quad & \exp\left\{\frac{(x^p - \mathrm{M}_p - y\Delta_p \mathrm{B}_p)^T \Delta_p^{-1}(x^p - \mathrm{M}_p - y\Delta_p \mathrm{B}_p) + \delta^2}{y} + y(\alpha^2 - \mathrm{B}_p^T \Delta_p \mathrm{B}_p)\right\} \\ \propto \quad & y^{\lambda-\frac{p}{2}-1} \exp\left\{\frac{(x^p - \mathrm{M}_p)^T \Delta_p^{-1}(x^p - \mathrm{M}_p) + \delta^2}{y} + y\alpha^2\right\}. \end{aligned}$$

Hence, normalizing the above quantity we have

$$f_{Y|X^p}(y \mid x^p) = \mathrm{GIG}\left(y; \lambda - \frac{p}{2}, {r_p}^2, \alpha^2\right),$$

where $r_p = \sqrt{(x^p - \mathrm{M}_p)^T \Delta_p^{-1}(x^p - \mathrm{M}_p) + \delta^2}$.

Using the updating mechanism (4.48), we can construct an i-order Markov transition probability as follows

$$\begin{aligned} p(x_{t+i} \mid x^{(t,i-1)}) &= \int_{\mathbb{R}_+} \mathrm{N}\left(x_{t+i}; \mu_{t+i} + y\beta_{t+i}, y\right) \mathrm{GIG}\left(y; \lambda - \frac{i}{2}, r^2_{(t,i-1)}, \alpha^2\right) dy \\ &= \mathrm{GH}\left(x_{t+i}; \lambda - \frac{i}{2}, \sqrt{\alpha^2 + \beta^2_{t+i}}, \beta_{t+i}, r_{(t,i-1)}, \mu_{t+i}\right). \end{aligned} \tag{4.62}$$

The integration in the above equation is easily found by multiplying the appropriate constant so that

$$\mathrm{GIG}\left(\lambda - \frac{i+1}{2}, (x_{t+i} - \mu_{t+i})^2 + r^2_{(t,i-1)}, \alpha^2 + \beta^2_{t+i}\right). \tag{4.63}$$

integrates out one. Here, we have treated the case where $|\Delta_p| = 1$, however the as-

sumption that $F_{X|Y}(\cdot, y) = \mathrm{N}(\mu_{t+i} + y\beta_{t+i}, y)$ in transition probability (4.62) imposes that the diagonal elements of Δ_p must be equal to one, that is the identity matrix, as stated in equation (4.59). With the stated generality for the matrix Δ and the vectors M and B, we can recover general lag-dependence models such as (4.58), where the lagged values enter with a different weight. On the other hand, with this generality for the i-order Markov transition probabilities (4.62), we do not necessarily have the existence of an invariant distribution for the p-lagged process X. However, under some restrictions on the parameters we can provide an equivalent to Proposition 4.4.

Proposition 4.5. Assume that $\Delta = \mathrm{I}$, $\mathrm{B} = (\beta, \beta, \ldots, \beta)^T$, $\beta_{t+i} = \beta$, $\mathrm{M} = (\mu, \mu, \ldots, \mu)^T$ and $\mu_{t+i} = \mu$, for $i = 1, \ldots, p$. If $X_t \sim \mathrm{GH}(\lambda, \alpha, \beta, \delta, \mu)$ and $X_{t+i} \mid X^{(t,i-1)}$ follows distribution (4.62) for $i = 1, \ldots, p$, then marginally $X_{t+p} \sim \mathrm{GH}(\lambda, \alpha, \beta, \delta, \mu)$.

Proof. The proof follows by the same argument used in the proof of Proposition 4.4. □

When conditions in Proposition 4.5 are satisfied the resulting ARCH model can be seen as a special case of the model proposed by Barndorff-Nielsen [9] when

$$r(X^{(t,i)}, \delta, \mu) = r_{(t,i)} = \left(\delta^2 + \sum_{j=0}^{i} (X_{t+j} - \mu)^2 \right)^{1/2}. \tag{4.64}$$

However, in [9] the existence of the invariant density is not imposed. In other words, the parameter restrictions stated in Proposition 4.5 can be thought as sufficient conditions for stationarity.

The stationary GH-ARCH(p) model can be written in the following from

$$X_{t+p} = \mu + \sqrt{\delta^2 + \sum_{j=1}^{p} (X_{t+p-j} - \mu)^2}\, \varepsilon_{t+p}, \tag{4.65}$$

where

$$\varepsilon_{t+p} \sim \mathrm{GH}\left(\lambda - \frac{p}{2}, \sqrt{\alpha^2 + \beta^2}, \beta, 1, 0\right).$$

One aspect of interest, for the sake of comparison with other ARCH models, is that the model (4.65) can be expressed in terms of a variable ε with zero mean and unit

variance. However, in the general case, for any choice of λ, this is neither analytically nor numerically simple. This because we need to deal with zeros of modified Bessel functions which depend on $r_{(\cdot,\cdot)}$, or more specifically on the lagged values. See Andersson [3] and Jensen and Lunde [47] for applications of non-stationary models when the mentioned standardization is simpler ($\lambda = -1/2$). In [3], the case when $\beta = 0$ and $\lambda = -1/2$ (NIG case) is studied and compared with some other ARCH models. In [47] a more complete comparison study is done, although the assumption of $\lambda = -1/2$ was still made.

The correlation structure for the stationary GH-ARCH(p) model gets more complicated than the one-lag case ($p = 1$) already presented in Section 4.4. However, in Figure 4.5 we have showed, based on simulated data via representation (4.48), how the autocorrelation of X behaves as the order changes. The higher the lag-dependence is the higher the correlation observed. Here, we can also observe that $\beta \neq 0$ allows for high (no-squared) autocorrelation.

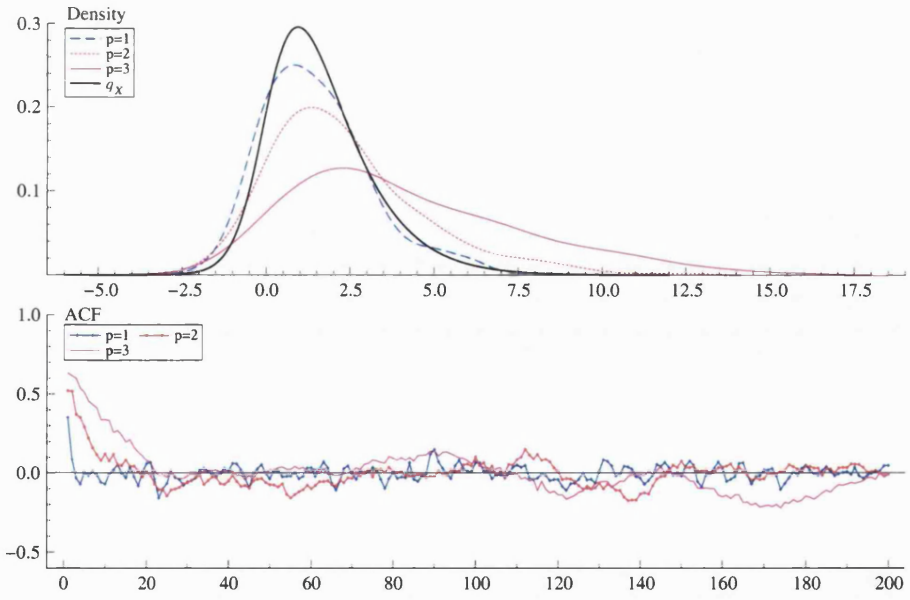


Figure 4.5: Densities for 500 simulated data from stationary GH-ARCH(p) models. The parameters are $\lambda = 2, \alpha = 2, \beta = 1, \delta = 1$ and $\mu = 0$. The density in bold represents the invariant distribution. For the simulated data, a burn in of 2000 simulations was applied.

4.8 PARAMETER ESTIMATION VIA MLE

In this section we carry out the parameter estimation via MLE, as described in Section 3.7, for the stationary GH-ARCH(p) model introduced in Section 4.7.

First, let us write the corresponding likelihood in terms of a p-lagged model and a given sample $\mathbf{x} = (x_1, x_2, \ldots, x_T)$

$$L_{\mathbf{x}}(\theta) = q_X^{\theta}(x_1) \left\{ \prod_{i=1}^{p-1} p(x_{i+1} \mid x^{(1,i-1)}) \right\} \left\{ \prod_{i=1}^{T-p} p(x_{i+p} \mid x^{(i,p-1)}) \right\}, \tag{4.66}$$

where $\theta = (\lambda, \alpha, \beta, \delta, \mu)$. In the case of a stationary GH-ARCH(p) model, the transition probabilities in (4.66) are given by (4.62).

The first step towards the maximum likelihood estimation is given by obtaining the gradient ∇l_θ, where $l_\theta = \log L_{\mathbf{x}}(\theta)$. To simplify notation let us assume that

$$\zeta = \sqrt{\alpha^2 + \beta^2}, \omega = \delta\sqrt{\alpha^2 - \beta^2}, \eta = \delta/\sqrt{\alpha^2 - \beta^2}$$

and $r_{(t,i)}$ is given as in equation (4.64). With this notation and the differentiation rules given in Table 4.1, direct differentiation leads to the following expressions for the components of the gradient ∇l_θ

$$\begin{aligned} \frac{\partial l_\theta}{\partial \alpha} &= \left\{ R_\lambda(\omega)\alpha\eta - R_{\lambda-\frac{1}{2}}(\alpha r_{(1,0)}) r_{(1,0)} \right\} \\ &+ \sum_{i=1}^{p-1} \left\{ R_{\lambda-\frac{i}{2}}(\alpha r_{(1,i-1)}) r_{(1,i-1)} - \frac{\alpha}{\zeta} R_{\lambda-\frac{i+1}{2}}(\zeta r_{(1,i)}) r_{(1,i)} \right\} \\ &+ \sum_{i=1}^{T-p} \left\{ R_{\lambda-\frac{p}{2}}(\alpha r_{(i,p-1)}) r_{(i,p-1)} - \frac{\alpha}{\zeta} R_{\lambda-\frac{p+1}{2}}(\zeta r_{(i,p)}) r_{(i,p)} \right\}, \end{aligned} \tag{4.67}$$

$$\begin{aligned} \frac{\partial l_\theta}{\partial \beta} &= \{(x_1 - \mu) - \beta\eta R_\lambda(\omega)\} \\ &+ \sum_{i=1}^{p-1} \left\{ (x_{i+1} - \mu) - \frac{\beta}{\zeta} R_{\lambda-\frac{i+1}{2}}(\zeta r_{(1,i)}) r_{(1,i)} \right\} \\ &+ \sum_{i=1}^{T-p} \left\{ (x_{i+p} - \mu) - \frac{\beta}{\zeta} R_{\lambda-\frac{p+1}{2}}(\zeta r_{(i,p)}) r_{(i,p)} \right\}, \end{aligned} \tag{4.68}$$

$$\frac{\partial l_\theta}{\partial \delta} = \left\{ \frac{2\delta\left(\lambda-\frac{1}{2}\right)}{r^2_{(1,0)}} - \frac{2\lambda}{\delta} + \frac{\omega}{\delta} R_\lambda(\omega) - \frac{\delta\alpha}{r_{(1,0)}} R_{\lambda-\frac{1}{2}}(\alpha r_{(1,0)}) \right\} \tag{4.69}$$
$$+ \sum_{i=1}^{p-1} \left\{ \frac{2\delta\left(\lambda-\frac{i+1}{2}\right)}{r^2_{(1,i)}} - \frac{2\delta\left(\lambda-\frac{i}{2}\right)}{r^2_{(1,i-1)}} + \frac{\alpha\delta R_{\lambda-\frac{i}{2}}(\alpha r_{(1,i-1)})}{r_{(1,i-1)}} - \frac{\zeta\delta R_{\lambda-\frac{i+1}{2}}(\zeta r_{(1,i)})}{r_{(1,i)}} \right\}$$
$$+ \sum_{i=1}^{T-p} \left\{ \frac{2\delta\left(\lambda-\frac{p+1}{2}\right)}{r^2_{(i,p)}} - \frac{2\delta\left(\lambda-\frac{p}{2}\right)}{r^2_{(i,p-1)}} + \frac{\alpha\delta R_{\lambda-\frac{p}{2}}(\alpha r_{(i,p-1)})}{r_{(i,p-1)}} - \frac{\zeta\delta R_{\lambda-\frac{p+1}{2}}(\zeta r_{(i,p)})}{r_{(i,p)}} \right\},$$

$$\frac{\partial l_\theta}{\partial \mu} = \left\{ \frac{\alpha(x_1-\mu)}{r_{(1,0)}} R_{\lambda-\frac{1}{2}}(\alpha r_{(1,0)}) - \frac{2\left(\lambda-\frac{1}{2}\right)(x_1-\mu)}{r^2_{(1,0)}} - \beta T \right\}$$
$$+ \left\{ \sum_{i=1}^{p-1} \frac{\mathbf{1}(x^{(1,i)}-\mu)}{r_{(1,i)}} \left(R_{\lambda-\frac{i+1}{2}}(\zeta r_{(1,i)})\zeta - \frac{2\left(\lambda-\frac{i+1}{2}\right)}{r_{(1,i)}} \right) \right.$$
$$\left. - \frac{\mathbf{1}(x^{(1,i-1)}-\mu)}{r_{(1,i-1)}} \left(R_{\lambda-\frac{i}{2}}(\alpha r_{(1,i-1)})\alpha - \frac{2\left(\lambda-\frac{i}{2}\right)}{r_{(1,i-1)}} \right) \right\}$$
$$+ \left\{ \sum_{i=1}^{T-p} \frac{\mathbf{1}(x^{(i,p)}-\mu)}{r_{(i,p)}} \left(R_{\lambda-\frac{p+1}{2}}(\zeta r_{(i,p)})\zeta - \frac{2\left(\lambda-\frac{p+1}{2}\right)}{r_{(i,p)}} \right) \right.$$
$$\left. - \frac{\mathbf{1}(x^{(i,p-1)}-\mu)}{r_{(i,p-1)}} \left(R_{\lambda-\frac{p}{2}}(\alpha r_{(i,p-1)})\alpha - \frac{2\left(\lambda-\frac{p}{2}\right)}{r_{(i,p-1)}} \right) \right\}, \tag{4.70}$$

where $\mathbf{1}(x^{(i,t)}-\mu) = \sum_{j=i}^{t+i}(x_j-\mu)$. Finally,

$$\frac{\partial l_\theta}{\partial \lambda} = \left\{ \log\left(\frac{r_{(1,0)}}{\alpha\eta}\right) - \frac{\dot{K}_{\lambda-\frac{1}{2}}(\alpha r_{(1,0)})}{K_{\lambda-\frac{1}{2}}(\alpha r_{(1,0)})} - \frac{\dot{K}_\lambda(\omega)}{K_\lambda(\omega)} \right\}$$
$$+ \sum_{i=1}^{p-1} \left\{ \log\left(\frac{\alpha r_{(1,i)}}{\zeta r_{(1,i-1)}}\right) + \frac{\dot{K}_{\lambda-\frac{i+1}{2}}(\zeta r_{(1,i)})}{K_{\lambda-\frac{i+1}{2}}(\zeta r_{(1,i)})} - \frac{\dot{K}_{\lambda-\frac{i}{2}}(\alpha r_{(1,i-1)})}{K_{\lambda-\frac{i}{2}}(\alpha r_{(1,i-1)})} \right\}$$
$$+ \sum_{i=1}^{T-p} \left\{ \log\left(\frac{\alpha r_{(i,p)}}{\zeta r_{(i,p-1)}}\right) + \frac{\dot{K}_{\lambda-\frac{p+1}{2}}(\zeta r_{(i,p)})}{K_{\lambda-\frac{p+1}{2}}(\zeta r_{(i,p)})} - \frac{\dot{K}_{\lambda-\frac{p}{2}}(\alpha r_{(i,p-1)})}{K_{\lambda-\frac{p}{2}}(\alpha r_{(i,p-1)})} \right\}, \tag{4.71}$$

where $\dot{K}_\nu(x) = \partial K_\nu/\partial\nu$. For the differentials of Bessel functions with respect to their order we refer to Abramowitz and Stegun [2] (expressions 9.6.43 and 9.6.45).

Analytical maximization for the complete set of parameters $\theta = (\alpha, \beta, \delta, \mu, \lambda)$ is not possible. One approach of computing the MLE estimators is given via numerical maximization. The structure of the problem at issue lead us to multi-dimensional procedures. Here we have chosen the Broyden-Fletcher-Goldfarb-Shanno (BFGS) algorithm to maximize the likelihood. A description of this algorithm together with its code in

C can be found in Press et al. [80]. The BFGS algorithm uses the gradient and the numerical Hessian. A problem while using the gradient given by equations (4.67-4.71) is the number of evaluations of Bessel functions required, this clearly leads to a speed problem. To overcome this, we have decided to use the numerical gradient only for the first iterations the analytic expressions. The estimation algorithm is stopped if the relative change for all the individual parameters is less or equal than 10^{-8}.

Another important point while implementing the BFGS algorithm, as in most Newton like routines, is the choice of starting points. There are many ways of providing such starting points, for instance, we could choose the parameters based on an IID sample. We tested the program **hyp** implemented by Blæsild and Sørensen [16] to obtain MLE of an IID sample from hyperbolic distributions[4] and then to provide starting points for all the parameters and $\lambda = 1$. However, this led to similar difficulties since IID estimates do not necessarily give good guesses for the initial values.

Example 4.3. In order to illustrate the estimation mechanism described above, we have simulated 2000 values of a stationary GH-ARCH(2) model with parameters $\lambda = 2, \alpha = 2, \beta = 1, \delta = 1$ and $\mu = 0$. Once this has been done, the BFGS algorithm needed 40 iterations to get the required accuracy (10^{-8}). In Table 4.2 we have presented the iterations results, notice that from iteration 35 the change in the log-likelihood values are of order less than 10^{-6}. For this particular example the estimation was relatively fast taking 14.38 seconds on a PC with a Pentium IV 2.2 GHz processor. However, the speed issue becomes a problem when the parameter values in the iterations get closer to their corresponding boundaries. For the above example we have only used the numerical gradient. Figure 4.6 shows the smoothed densities corresponding to the parameter estimates of 500 sample paths. Notice how the modes concentrate around the true values. ○

4.9 PARAMETER ESTIMATION VIA EM

We have seen in Section 3.7 that an alternative method to estimate the parameters of a model constructed via the Gibbs sampler type scheme can be done using the

[4]We thank Prof. Sørensen for kindly providing us with their code. With this code it is possible to estimate the parameters of a hyperbolic distribution, that is GH$(1, \alpha, \beta, \delta, \mu)$, under an IID sampling scheme.

Iterations	λ	α	β	δ	μ	l_θ
(Model)	2.00000	2.00000	1.00000	1.00000	0.00000	-3740.490000
(Initial)	3.00000	1.50000	1.30000	3.00000	1.00000	-5660.629660
5	5.9883	3.0955	1.1508	4.3038	-0.51770	-3819.026506
10	6.5386	3.1259	1.6052	1.0925	-0.73760	-3744.320250
15	2.6838	2.2095	1.2552	1.3056	-0.30540	-3722.596794
20	2.1124	1.7577	0.94377	0.95160	0.12031	-3716.275118
25	2.0136	1.7883	0.97138	1.1200	0.063707	-3715.906539
30	1.6197	1.7397	0.97812	1.3626	0.052813	-3715.820703
35	1.6243	1.7399	0.97795	1.3595	0.053192	-3715.820677
40	1.6243	1.7399	0.97795	1.3595	0.053192	-3715.820677

Table 4.2: Iteration results of the BFGS algorithm corresponding to the MLE method in Example 4.3.

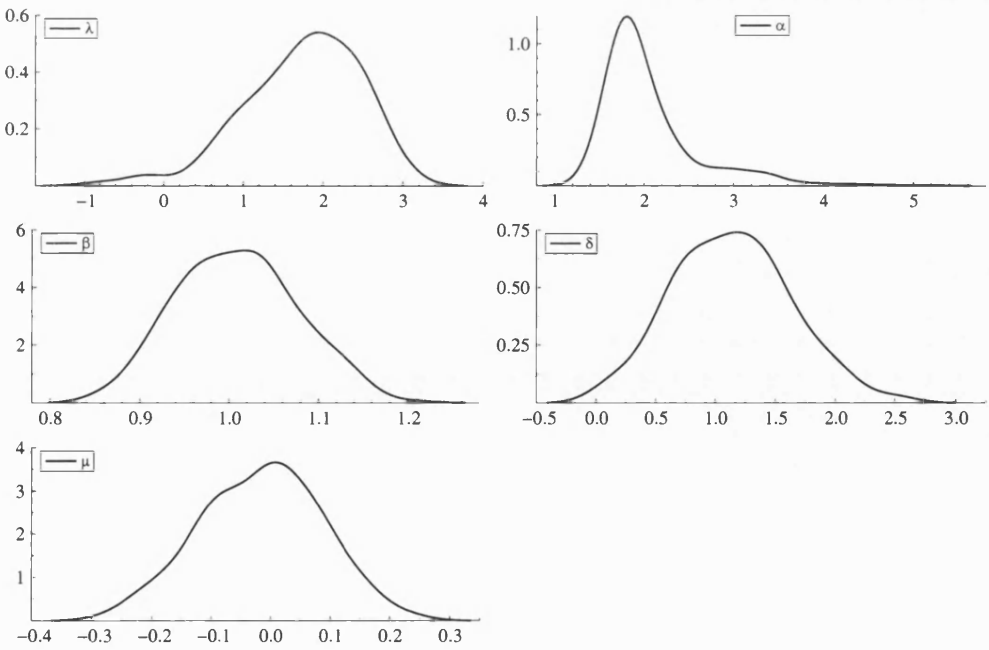


Figure 4.6: Smoothed densities for the parameter estimates corresponding to 500 sample paths (with 2000 simulations) of a GH-ARCH(2) model with parameters $\lambda = 2, \alpha = 2, \beta = 1, \delta = 1$ and $\mu = 0$.

EM algorithm. In this section we give the details of this estimation method for the stationary GH-ARCH(p) model.

The EM algorithm consists of iteratively computing the expectation and maximization steps given by

$$Q(\theta \mid \theta_{(j)}, \mathbf{x}) = \mathbb{E}_{\theta_{(j)}} \left[\log L^{aug}_{\mathbf{x},\mathbf{y}}(\theta) \right] \quad \text{and} \quad \theta_{(j+1)} = \arg\max_{\theta} Q(\theta \mid \theta_{(j)}, \mathbf{x}), \tag{4.72}$$

respectively. Here $L^{aug}_{\mathbf{x},\mathbf{y}}(\theta)$ denotes the augmented likelihood and $\theta_{(j)}$, $j = 1, 2, \ldots$ stands for the current parameter value (or initial value if $j = 0$). The expectation $\mathbb{E}_{\theta_{(j)}}[\cdot]$ is taken with respect to $\mathrm{F}^{\theta_{(j)}}_{\mathbf{Y}|\mathbf{X}}$, that is, the distribution of the latent information given the observations and the current parameter values.

In this case the required augmented likelihood (3.39) is given by

$$\begin{aligned} L^{aug}_{\mathbf{x},\mathbf{y}}(\theta) &= q^{\theta}_{X}(x_1) \prod_{i=1}^{p-1} f^{\theta}_{X|Y}(x_{i+1} \mid y_{i+1})\, f^{\theta}_{Y|X}(y_{i+1} \mid x^{(1,i-1)}) \\ &\times \prod_{i=1}^{T-p} f^{\theta}_{X|Y}(x_{i+p} \mid y_{i+p})\, f^{\theta}_{Y|X}(y_{i+p} \mid x^{(i,p-1)}). \end{aligned}$$

Therefore, for the GH-ARCH(p) case, we have

$$\begin{aligned} l^{aug}_{\mathbf{x},\mathbf{y}}(\theta) &= \log\left(\mathrm{GH}(x_1, \lambda, \alpha, \beta, \delta, \mu)\right) + \sum_{i=1}^{T-1} \log\left(\mathrm{N}\left(x_{i+1}; \mu + \beta y_{i+1}, y_{i+1}\right)\right) \\ &+ \sum_{i=1}^{p-1} \log\left(\mathrm{GIG}\left(y_{i+1}; \lambda - \frac{i}{2}, r^2_{(1,i-1)}, \alpha^2\right)\right) \\ &+ \sum_{i=1}^{T-p} \log\left(\mathrm{GIG}\left(y_{i+p}; \lambda - \frac{p}{2}, r^2_{(i,p-1)}, \alpha^2\right)\right), \end{aligned} \tag{4.73}$$

where $l^{aug}_{\mathbf{x},\mathbf{y}}(\theta) = \log L^{aug}_{\mathbf{x},\mathbf{y}}(\theta)$.

For the augmented likelihood, the gradient simplifies slightly with respect to gradient of the non-augmented likelihood. Using the same notation as in (4.67-4.71) we get the following expressions

$$\frac{\partial l^{aug}_{\mathbf{x},\mathbf{y}}(\theta)}{\partial \alpha} = \left\{R_\lambda(\omega)\alpha\eta - R_{\lambda-\frac{1}{2}}(\alpha r_{(1,0)})r_{(1,0)}\right\} + \sum_{i=1}^{p-1}\left\{R_{\lambda-\frac{i}{2}}(\alpha r_{(1,i-1)})r_{(1,i-1)}\right\} + \sum_{i=1}^{T-p}\left\{R_{\lambda-\frac{p}{2}}(\alpha r_{(i,p-1)})r_{(i,p-1)}\right\} - \alpha\sum_{i=1}^{T-1} y_{i+1}, \tag{4.74}$$

$$\frac{\partial l^{aug}_{\mathbf{x},\mathbf{y}}(\theta)}{\partial \beta} = \sum_{i=1}^{T} x_i - \beta\sum_{i=1}^{T-1} y_{i+1} - T\mu - \beta\eta R_\lambda(\omega), \tag{4.75}$$

$$\begin{aligned}\frac{\partial l^{aug}_{\mathbf{x},\mathbf{y}}(\theta)}{\partial \delta} &= \left\{\frac{2\delta\left(\lambda-\frac{1}{2}\right)}{r^2_{(1,0)}} - \frac{2\lambda}{\delta} + \frac{\omega}{\delta}R_\lambda(\omega) - \frac{\delta\alpha}{r_{(1,0)}}R_{\lambda-\frac{1}{2}}(\alpha r_{(1,0)})\right\} \\ &+ \sum_{i=1}^{p-1}\left\{R_{\lambda-\frac{i}{2}}(\alpha r_{(1,i-1)})\frac{\delta\alpha}{r_{(1,i-1)}} - \frac{2\delta\left(\lambda-\frac{i}{2}\right)}{r^2_{(1,i-1)}}\right\} \\ &+ \sum_{i=1}^{T-p}\left\{R_{\lambda-\frac{p}{2}}(\alpha r_{(i,p-1)})\frac{\delta\alpha}{r_{(i,p-1)}} - \frac{2\delta\left(\lambda-\frac{p}{2}\right)}{r^2_{(i,p-1)}}\right\} - \delta\sum_{i=1}^{T-1}\frac{1}{y_{i+1}},\end{aligned} \tag{4.76}$$

$$\begin{aligned}\frac{\partial l^{aug}_{\mathbf{x},\mathbf{y}}(\theta)}{\partial \mu} &= R_{\lambda-\frac{1}{2}}(\alpha r_{(1,0)})\frac{(x_1-\mu)\alpha}{r_{(1,0)}} - \beta T - \frac{2\left(\lambda-\frac{1}{2}\right)(x_1-\mu)}{r^2_{(1,0)}} + \sum_{i=1}^{T-1}\frac{x_{i+1}-\mu}{y_{i+1}} \\ &+ \sum_{i=1}^{p-1}\mathbf{1}(x^{(1,i-1)}-\mu)\left\{\frac{2\left(\lambda-\frac{i}{2}\right)}{r^2_{(1,i-1)}} + \frac{1}{y_{i+1}} - \frac{R_{\lambda-\frac{i}{2}}(\alpha r_{(1,i-1)})\alpha}{r_{(1,i-1)}}\right\} \\ &+ \sum_{i=1}^{T-p}\mathbf{1}(x^{(i,p-1)}-\mu)\left\{\frac{2\left(\lambda-\frac{p}{2}\right)}{r^2_{(i,p-1)}} + \frac{1}{y_{i+p}} - \frac{R_{\lambda-\frac{p}{2}}(\alpha r_{(i,p-1)})\alpha}{r_{(i,p-1)}}\right\},\end{aligned} \tag{4.77}$$

$$\begin{aligned}\frac{\partial l^{aug}_{\mathbf{x},\mathbf{y}}(\theta)}{\partial \lambda} &= \left\{\log\left[\frac{r_{(1,0)}}{\alpha\eta}\right] - \frac{\dot{K}_{\lambda-\frac{1}{2}}(\alpha r_{(1,0)})}{K_{\lambda-\frac{1}{2}}(\alpha r_{(1,0)})} - \frac{\dot{K}_\lambda(\omega)}{K_\lambda(\omega)}\right\} + \sum_{i=1}^{T-1}\log(y_{i+1}) \\ &+ \sum_{i=1}^{p-1}\left\{\log\left[\frac{\alpha}{r_{(1,i-1)}}\right] - \frac{\dot{K}_{\lambda-\frac{i}{2}}(\alpha r_{(1,i-1)})}{K_{\lambda-\frac{i}{2}}(\alpha r_{(1,i-1)})}\right\} \\ &+ \sum_{i=1}^{T-p}\left\{\log\left[\frac{\alpha}{r_{(i,p-1)}}\right] - \frac{\dot{K}_{\lambda-\frac{p}{2}}(\alpha r_{(i,p-1)})}{K_{\lambda-\frac{p}{2}}(\alpha r_{(i,p-1)})}\right\}.\end{aligned} \tag{4.78}$$

The main difficulty with the EM algorithm is to obtain the expectation step, since the distribution $F^{\theta(j)}_{\mathbf{Y}|\mathbf{X}}$ might not have a simple form. As we mentioned before, in the case of

models constructed via the Gibbs sampler type scheme, we can depict the distribution of $\mathrm{F}^{\theta}_{\mathbf{Y}|\mathbf{X}}$ by considering component-wise the conditional independent variables with density function

$$f(y_{t+i} \mid x_{t+i}, x^{(t,i-1)}) \propto f^{\theta}_{X|Y}(x_{t+i} \mid y_{t+i})\, f^{\theta}_{Y|X}(y_{t+i} \mid x^{(t,i-1)}).$$

Where $x^{(t,i-1)} = (x_t, x_{t+i,\ldots,x_{t+i-1}})$. In this case we have the following result

$$\begin{aligned} f(y_{t+i} \mid x_{t+i}, x^{(t,i-1)}) &= \mathrm{GIG}\left(y_{t+i}; \lambda - \frac{i+1}{2}, (x_{t+i}-\mu)^2 + r^2_{(t,i-1)}, \alpha^2+\beta^2\right) \\ &= \mathrm{GIG}\left(y_{t+i}; \lambda - \frac{i+1}{2}, r^2_{(t,i)}, \alpha^2+\beta^2\right). \end{aligned}$$

for $i = 1, \ldots, T-1$. Which is precisely the same as equation (4.63). Therefore, for a stationary GH-ARCH(p) model, we can depict the random vector $\{\mathbf{Y} \mid \mathbf{X}\}$ by considering the following conditionally independent random variables

$$\begin{aligned} Y_1 &\sim \mathrm{GIG}(\lambda, \delta^2, \alpha^2-\beta^2), \\ Y_{i+1} \mid x^{(1,i)} &\sim \mathrm{GIG}\left(\lambda - \frac{i+1}{2}, r^2_{(1,i)}, \alpha^2+\beta^2\right), \quad \text{for } i = 1, \ldots, p-1 \\ Y_{i+p} \mid x^{(i,p)} &\sim \mathrm{GIG}\left(\lambda - \frac{p+1}{2}, r^2_{(i,p)}, \alpha^2+\beta^2\right), \quad \text{for } i = 1, \ldots, T-p. \end{aligned} \tag{4.79}$$

Using this decomposition, it is possible to compute the density $f^{\theta_{(j)}}_{\mathbf{Y}|\mathbf{X}}$ for the required distribution by simply multiplying the densities corresponding to the above random variables. This decomposition is also useful to simulate from $\mathbf{Y} \mid \mathbf{X}$, which can be used in the MCEM scheme described in next section.

In Walker [100], it was shown that the M-step involved in the EM algorithm can be simplified as follows; if we let $\theta^{-i}_{(j)} = (\theta^1_{(j)}, \ldots, \theta^{i-1}_{(j)}, \theta^{i+1}_{(j)}, \ldots, \theta^d_{(j)})$, hence the M-step can be simplified by component-wise solving

$$\frac{\partial Q(\theta^i \mid \theta^{-i}_{(j)}, \mathbf{x})}{\partial \theta^i} = \mathbb{E}_{\theta_{(j)}}\left[\frac{\partial l^{aug}_{\mathbf{x},\mathbf{y}}(\theta^i)}{\partial \theta^i}\right]\Big|_{\theta^i = \theta^i_{(j+1)}} = 0 \tag{4.80}$$

for $i = 1, \ldots, d$ and where the expectation $\mathbb{E}_{\theta_{(j)}}[\cdot]$ is taken with respect to $\mathrm{F}^{\theta_{(j)}}_{\mathbf{Y}|\mathbf{X}}$. See also Louis [67].

In Walker [100], the expectation in (4.80) was computed with Monte Carlo simulations

from $\mathbf{Y} \mid \mathbf{X}$. In this case such expectation can be taken analytically by using decomposition (4.79). The following result will allow us to estimate the parameters of the stationary GH-ARCH(p) via a sequential MLE.

Theorem 4.2. Let $\mathbf{x} = (x_1, x_2, \ldots, x_T)$ be a sample from a stationary GH-ARCH(p) model, then

$$\mathbb{E}_{\theta_{(j)}}\left[\nabla l^{aug}_{\mathbf{x},\mathbf{y}}(\theta)\right] = \nabla l_{\mathbf{x}}(\theta). \tag{4.81}$$

Proof. All we need is to consider the first moments for the sufficient statistics involved in the augmented gradient $l^{aug}_{\mathbf{x},\mathbf{y}}$. First notice that if $Y \sim \text{GIG}(\lambda, \delta, \gamma)$ and $Z = 1/Y$ then $Z \sim \text{GIG}(-\lambda, \gamma, \delta)$. Therefore, using (4.20), we see that

$$\mathbb{E}\left[\frac{1}{Y}\right] = \frac{R_{-\lambda}(\omega)}{\eta},$$

where $\omega = \sqrt{\delta\gamma}$ and $\eta = \sqrt{\delta/\gamma}$. Now, using property 4 in Table 4.1, we see that

$$R_{-\lambda}(\omega) = R_{\lambda}(\omega) - \frac{2\lambda}{\omega}.$$

Hence,

$$\mathbb{E}\left[\frac{1}{Y}\right] = \frac{1}{\eta}\left\{R_{\lambda}(\omega) - \frac{2\lambda}{\omega}\right\}.$$

Applying the above result to the corresponding GIG distributions involved in (4.79) we get

$$\mathbb{E}\left[\frac{1}{Y_{i+1}}\right] = \frac{\zeta R_{\lambda-\frac{i+1}{2}}(r_{(1,i)}\zeta)}{r_{(1,i)}} - \frac{r^2_{(1,i)}\lambda - \frac{i+1}{2}\big)}{r^2_{(1,i)}}, \tag{4.82}$$

provided that $Y_{i+1} \sim \text{GIG}\left(\lambda - \frac{i+1}{2}, r^2_{(1,i)}, \alpha^2 + \beta^2\right)$, and

$$\mathbb{E}\left[\frac{1}{Y_{i+p}}\right] = \frac{\zeta R_{\lambda-\frac{p+1}{2}}(r_{(i,p)}\zeta)}{r_{(i,p)}} - \frac{2\left(\lambda - \frac{p+1}{2}\right)}{r^2_{(i,p)}}, \tag{4.83}$$

provided that $Y_{i+p} \sim \text{GIG}\left(\lambda - \frac{p+1}{2}, r^2_{(i,p)}, \alpha^2 + \beta^2\right)$.

On the other hand, if $Z = \log Y$ then

$$f_Z(\xi) = \mathrm{GIG}(e^{\xi}; \lambda, \delta, \gamma)\, e^{\xi}, \quad \xi \in \mathbb{R}.$$

With this the Laplace transform for Z is given by

$$\mathcal{L}_Z(\kappa) = \frac{K_{\lambda+\kappa}(\omega)}{K_\lambda(\omega)} \eta^\kappa$$

and so

$$\begin{aligned} \mathbb{E}[\log(Y)] &= \mathcal{L}_Z'(0) \\ &= \frac{\dot{K}_\lambda(\omega)}{K_\lambda(\omega)} + \log(\eta), \end{aligned} \tag{4.84}$$

where $\dot{K}_\nu(x) = \partial K_\nu(x)/\partial \nu$.

Therefore, applying the above result as in (4.82) and (4.83) we get

$$\mathbb{E}[\log(Y_{i+1})] = \frac{\dot{K}_{\lambda-\frac{i+1}{2}}(r_{(1,i)}\zeta)}{K_{\lambda-\frac{i+1}{2}}(r_{(1,i)}\zeta)} + \log\left(\frac{r_{(1,i)}}{\zeta}\right) \tag{4.85}$$

and

$$\mathbb{E}[\log(Y_{i+p})] = \frac{\dot{K}_{\lambda-\frac{p+1}{2}}(r_{(i,p)}\zeta)}{K_{\lambda-\frac{p+1}{2}}(r_{(i,p)}\zeta)} + \log\left(\frac{r_{(i,p)}}{\zeta}\right) \tag{4.86}$$

respectively. In the same way, using (4.20), we compute

$$\mathbb{E}[Y_{i+1}] = R_{\lambda-\frac{i+1}{2}}(r_{(1,i)}\zeta)\,\frac{r_{(1,i)}}{\zeta}, \quad i = 1, \ldots, p-1 \tag{4.87}$$

and

$$\mathbb{E}[Y_{i+p}] = R_{\lambda-\frac{p+1}{2}}(r_{(i,p)}\zeta)\,\frac{r_{(i,p)}}{\zeta}, \quad i = 1, \ldots, T-p. \tag{4.88}$$

With the above moments, all the quantities required to compute the expectation $\mathbb{E}_{\theta^{(j)}}[\nabla l^{aug}_{\mathbf{x},\mathbf{y}}(\theta)]$ are provided.

For example

$$\begin{aligned}
\mathbb{E}_{\theta_{(j)}}\left[\frac{\partial l^{aug}_{\mathbf{x},\mathbf{y}}(\theta)}{\partial\lambda} \mid \mathbf{x}\right] &= \left\{\log\left(\frac{r_{(1,0)}}{\alpha\eta}\right) - \frac{\dot{K}_{\lambda-\frac{1}{2}}(\alpha r_{(1,0)})}{K_{\lambda-\frac{1}{2}}(\alpha r_{(1,0)})} - \frac{\dot{K}_{\lambda}(\omega)}{K_{\lambda}(\omega)}\right\} \\
&+ \sum_{i=1}^{p-1}\left\{\log\left(\frac{\alpha}{r_{(1,i-1)}}\right) - \frac{\dot{K}_{\lambda-\frac{i}{2}}(\alpha r_{(1,i-1)})}{K_{\lambda-\frac{i}{2}}(\alpha r_{(1,i-1)})}\right\} \\
&+ \sum_{i=1}^{T-p}\left\{\log\left(\frac{\alpha}{r_{(i,p-1)}}\right) - \frac{\dot{K}_{\lambda-\frac{p}{2}}(\alpha r_{(i,p-1)})}{K_{\lambda-\frac{p}{2}}(\alpha r_{(i,p-1)})}\right\} \\
&+ \mathbb{E}_{\theta_{(j)}}\left[\sum_{i=1}^{T-1}\log\left(Y_{i+1}\right) \mid \mathbf{x}\right],
\end{aligned}$$

where

$$\begin{aligned}
\mathbb{E}_{\theta_{(j)}}\left[\sum_{i=1}^{T-1}\log\left(Y_{i+1}\right) \mid \mathbf{x}\right] &= \sum_{i=1}^{p-1}\mathbb{E}^{\diamond}_{\theta_{(j)}}\left[\log\left(Y_{i+1}\right)\right] + \sum_{i=1}^{T-p}\mathbb{E}^{*}_{\theta_{(j)}}\left[\log\left(Y_{i+p}\right)\right] \\
&= \sum_{i=1}^{p-1}\left\{\frac{\dot{K}_{\lambda-\frac{i+1}{2}}(r_{(1,i)}\zeta)}{K_{\lambda-\frac{i+1}{2}}(r_{(1,i)}\zeta)} + \log\left(\frac{r_{(1,i)}}{\zeta}\right)\right\} \\
&+ \sum_{i=1}^{T-p}\left\{\frac{\dot{K}_{\lambda-\frac{p+1}{2}}(r_{(i,p)}\zeta)}{K_{\lambda-\frac{p+1}{2}}(r_{(i,p)}\zeta)} + \log\left(\frac{r_{(i,p)}}{\zeta}\right)\right\}.
\end{aligned}$$

In the first equality the expectations $\mathbb{E}^{\diamond}$ and $\mathbb{E}^{*}$ are given by (4.85) and (4.86) respectively. Re-arranging expressions we get equation (4.71). Hence, by simple substituting the required moments while taking the expectations of (4.74-4.78), we get all the others equalities in the same way and therefore the stated result (4.81) follows. □

Theorem 4.2 allows us to estimate the parameters of the stationary GH-ARCH(p) model as follows: Given a set of initial values $\theta_0 = (\lambda_0, \alpha_0, \beta_0, \delta_0, \mu_0)$, the first update (first iteration of the EM algorithm) is given by individually solving

$$\begin{aligned}
&\frac{\partial l_{\mathbf{x}}(\lambda,\alpha_0,\beta_0,\delta_0,\mu_0)}{\partial\lambda}\Big|_{\lambda=\lambda_1} = 0, \qquad \frac{\partial l_{\mathbf{x}}(\lambda_1,\alpha,\beta_0,\delta_0,\mu_0)}{\partial\alpha}\Big|_{\alpha=\alpha_1} = 0 \\
&\frac{\partial l_{\mathbf{x}}(\lambda_1,\alpha_1,\beta,\delta_0,\mu_0)}{\partial\beta}\Big|_{\beta=\beta_1} = 0, \qquad \frac{\partial l_{\mathbf{x}}(\lambda_1,\alpha_1,\beta_1,\delta,\mu_0)}{\partial\delta}\Big|_{\delta=\delta_1} = 0 \\
&\frac{\partial l_{\mathbf{x}}(\lambda_1,\alpha_1,\beta_1,\delta_1,\mu)}{\partial\mu}\Big|_{\mu=\mu_1} = 0.
\end{aligned} \tag{4.89}$$

Hence, we use θ_1 to get θ_2 and so on, until a required convergence criterion is satisfied. Ideally we would like equations (4.89) to have an analytical solution. For the general

Iterations	λ	α	β	δ	μ	l_θ
(Model)	2.00000	2.00000	1.00000	1.00000	0.00000	-3740.490000
(Initial)	3.00000	1.50000	1.30000	3.00000	1.00000	-5660.629660
50	-0.2627527	1.517131	1.007450	2.257465	0.000649	-3717.160942
100	0.7007581	1.627385	0.990147	1.828023	0.032054	-3716.186550
300	1.587935	1.734899	0.978064	1.378675	0.052922	-3715.821308
500	1.622195	1.739547	0.977901	1.360448	0.053250	-3715.820679
800	1.623464	1.739720	0.977894	1.359774	0.053263	-3715.820678
1000	1.623473	1.739723	0.977895	1.359771	0.053262	-3715.820678
1280	1.623497	1.739730	0.977897	1.359765	0.053260	-3715.820678
1300	1.623519	1.739737	0.977901	1.359760	0.053255	-3715.820677

Table 4.3: Iteration results of the EM algorithm corresponding to Example 4.4.

parameter domain of stationary GH-ARCH(p) this is not possible. However, numerical solutions for the above equations are much cheaper, in terms of time and efficiency, than the joint maximization procedure for the complete likelihood, where the Hessian is needed.

Notice that in general the above sequential MLE approach does not guaranty the convergence to the optimum value. However, in this case, Theorem 4.2 uses the EM algorithm to ensure such convergence. Some authors have found, based on empirical results, that this sequential MLE approach is useful, however, to the best of our knowledge, a general way to justify such approach is not available.

Example 4.4. For the sake of comparison we use the same model as the one used in Example 4.3. We also use the same convergence criterion. Therefore applying the EM algorithm, with the M-step as described in (4.89), we obtain the results displayed in Table 4.3. The non-linear equations underlying to (4.89) require the computation of modified Bessel functions, in this case we have used the routines described in Press et al. [80] to compute them. The number of EM iterations is considerably higher than those corresponding to the BFGS algorithm, however these iterations are much more faster. The time required to get an overall relative error less or equal than 10^{-6} was 38.45 sec. Also notice that, the convergence of the EM algorithm slows down when the estimates get closer to the optimum value. A way to improve this, in speed terms, is to switch to the MLE method after some initial EM iterations. ○

4.10 PARAMETER ESTIMATION VIA MCEM

As we have seen in Section 3.7 an alternative way of obtaining the expectation step in the EM algorithm is via the MCEM algorithm. That is, using Monte Carlo integration to approximate $Q(\theta \mid \theta_{(j)}, \mathbf{x})$ in (4.72). This approximation is given by

$$\hat{Q}(\theta \mid \theta_{(j)}, \mathbf{x}) = \frac{1}{m} \sum_{k=1}^{m} \log(L^{aug}_{\mathbf{x},\mathbf{y}^{(k)}}(\theta)), \tag{4.90}$$

where $\mathbf{y}^{(k)} \sim \mathrm{F}^{\theta_{(j)}}_{\mathbf{Y}|\mathbf{X}}$, $k = 1, \ldots, m$. In other words, we need to simulate m vectors $\mathbf{y}^{(k)}$, where each vector $\mathbf{y}^{(k)} = (y_{k,1}, y_{k,2}, \ldots, y_{k,T})$ can be simulated by individually simulating $y_{k,i}$ from distributions (4.79).

With the above approximation the maximization step involved in the MCEM method consists of maximizing (4.90) for which, in particular, the following gradient is needed

$$\nabla \widehat{Q}(\theta \mid \theta_{(j)}, \mathbf{x}) = \frac{1}{m} \sum_{k}^{m} \nabla l^{aug}_{\mathbf{x},\mathbf{y}^{(k)}}(\theta),$$

where the components of $\nabla l^{aug}_{\mathbf{x},\mathbf{y}^{(k)}}(\theta)$ are given by expressions (4.74-4.78). Even though in this case the analytic expression for this gradient is slightly simpler than for the MLE case, an analytic solution to the maximization problem is still not available. However, the same numeric approach taken in order to obtain the MLE, can be applied to maximize $\hat{Q}(\theta \mid \theta_{(j)}, \mathbf{x})$, namely the BFGS algorithm.

The contribution of the random vectors $\mathbf{y}^{(k)}$ to the partial differentials corresponding to expressions (4.74-4.78) is given through sufficient statistics. Therefore, the computationally "expensive" quantities (those involving Bessel functions) in the gradient are fixed within the MCEM iterations.

Here, it is worth mentioning that for some particular cases corresponding to GH-ARCH(p) models, those corresponding to sub-classes of GH distributions, both the MLE and the MCEM estimation methods are simplified. However, having decided to take the complete general approach, that is regarding all the parameters as unknown, then the MCEM method tends to get slower than the MLE as the sample size increases. This is, mainly caused by the fact that the lag-dependence functions $r_{(\cdot,\cdot)}$ tend to get bigger as the sample size increases. Therefore, the simulation from GIG distributions,

required in the approximation to the E-step in the MCEM algorithm, slows down[5].

4.11 REAL DATA EXAMPLE

On an empirical standpoint, it is known that time-varying volatility is well captured by ARCH-type models. See Engle [32]. However, it has been also accepted that in order to have better fitting ARCH-type models should be extended to include asymmetric and thick tailed distributions, as well as long-term dependence and cross correlation, see Gallant et al. [38]. Generalized hyperbolic distributions allow for skewness and thicker tails, therefore it seems natural to think in the stationary GH-ARCH model as a good candidate. Within the ARCH literature, the NIG subclass (non-stationary case) has been explored and proved to give better fit than Gaussian-based ARCH models. See Jensen and Lunde [47]. See also Barndorff-Nielsen [9] where the general exposition (without the stationarity assumptions) of GH-ARCH was introduced. As we mentioned before, the approach taken in [9] allows for a more general lag-dependence function (4.58). A more relevant specification for the function $r(\cdot,\cdot)$, from the stochastic volatility modelling point of view, is studied in [47]. However, due to the larger number of parameters in this specification, the estimation procedures become slower. Hence, the main objective of this section is to illustrate how, when working with the general stationary GH-ARCH(p) model, good results can be achieved with real financial data.

Another important point to underline here is that, due to the stationarity of our GH-ARCH model, we have Theorem 4.2, and therefore the estimation procedure simplifies considerably.

For this illustration we consider the following data sets:

Name	**Symbol**	**From**	**To**	**QLB**(30)
NASDAQ-100 SHARES	QQQ	01-Aug-99	01-Aug-01	605.51
NOKIA CORPORATION	NOK	01-Aug-98	01-Aug-01	1030.82
INTL BUS MACHINE	IBM	31-Jul-98	31-Jul-01	1047.08
DELL COMPUTER CORP	DELL	01-Aug-98	01-Aug-01	1045.12

[5]We use OX program language for simulating GIG distributions. See Doornik [28]. However, an alternative approach can be done using the methodology described in Chapter 5, see Example 5.3.

In the above table we have also displayed the Ljung-Box test for no autocorrelation in the squared returns (with 30 lags). In all cases, the presence of serial correlation was significant. The data series are showed in Figure 4.7. We assumed that the associated log-returns follow a stationary GH-ARCH(p) model. We tested models with order $p = 1, 2, 4, 8, 16$.

In Table 4.4 we present the parameter estimates based on the following: from the outset we consider 200 iterations of the EM algorithm described in Section 4.9. Once a good approximation is achieved, we switch to the BFGS algorithm with the numerical gradient described in Section 4.8. The last 15 trading days from each data set were left out from the historical period considered in the estimation to validate the forecasted values. See Figure 4.9. The absolute error used for the convergence was set to 10^{-6}. Notice that in all cases $\beta < 0$, which corroborates the empirical fact that log-returns are negatively correlated. This can also be seen in Figure 4.8 for the case of Nokia log-returns, where the density is slightly skewed to the right. In Table 4.4 we have also displayed Akaike's information criteria, notice that the models with minimum information tend to give better forecasts, as expected. From the ACF in Figure 4.8 we can see that the data shows significant correlation with the first and fourth squared-lagged-value. In particular we can also see that a GH-ARCH(4) allows for more correlation and the corresponding density seems to fit more adequately.

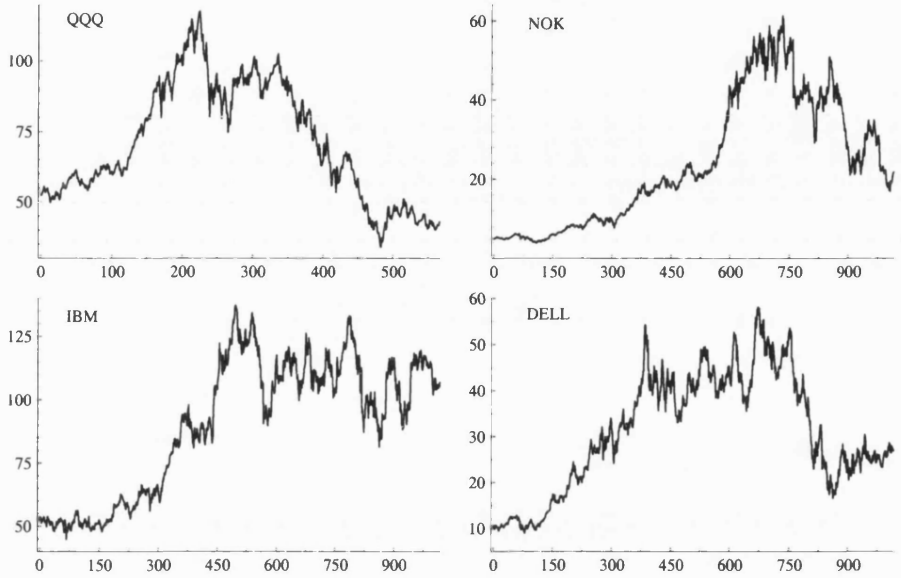


Figure 4.7: Data series for the periods: (01-Aug-99/01-Aug-01) for QQQ; (01-Aug-98/01-Aug-01) for NOK; (31-Jul-98/31-Jul-01) and (01-Aug-98/01-Aug-01) for DELL.

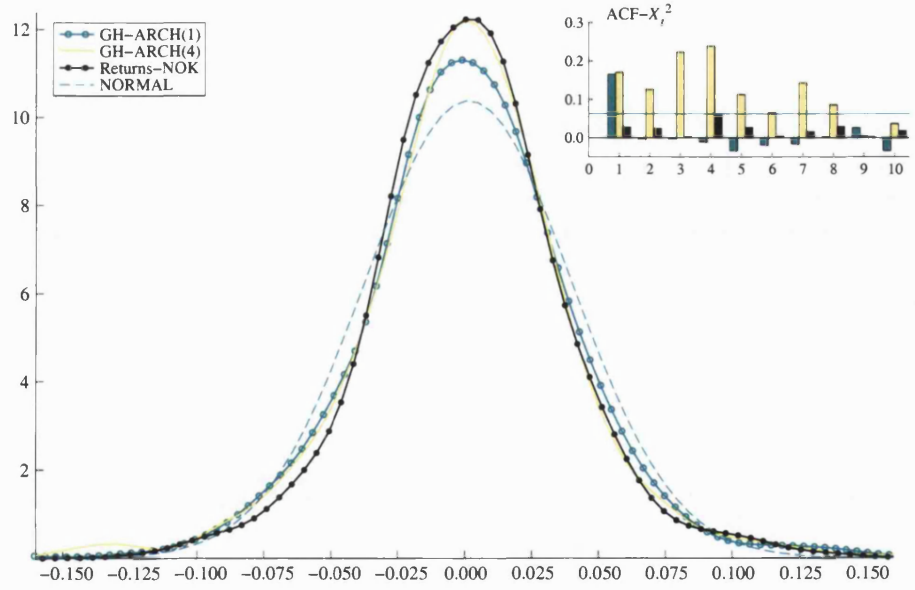


Figure 4.8: Density estimates and ACF's for Nokia daily log-returns and fitted stationary GH-ARCH(1) and GH-ARCH(4) models. The period for the data is: from 01-Aug-1998 to 01-Aug-2001. The underlying parameters are presented in Table 4.4.

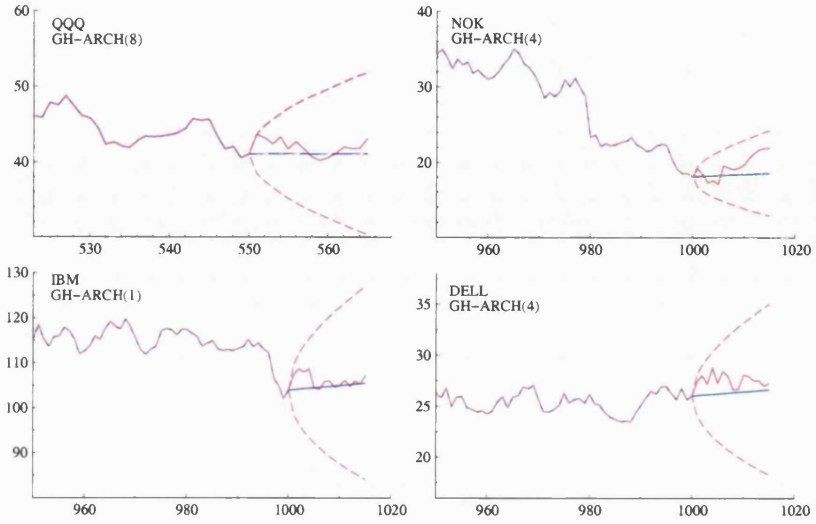


Figure 4.9: Forecasting corresponding to the models minimizing Akaike's criteria from Table 4.4.

Lag	λ	α	β	δ	μ	*loglik*	AIC	MC price
QQQ		QLB(30)	=605.51				Real price	=43.1
1	2.341	69.195	-6.084	0.015	0.006	1115.41	-2220.83	41.39
2	1.589	58.816	-7.490	0.014	0.007	1129.83	-2249.68	41.45
4	1.661	58.152	-6.270	0.008	0.006	1141.29	-2272.59	41.59
8	-0.043	43.500	-4.468	0.036	0.004	1147.00	**-2284.01**	41.59
16	-0.361	47.201	-2.979	0.047	0.003	1145.40	-2280.8	41.39
NOK		QLB(30)	=1030.82				Real price	=21.86
1	-2.116	7.270	-0.583	0.060	0.002	1914.01	-3818.02	19.57
2	-2.066	11.240	-2.828	0.061	0.005	1920.14	-3830.29	19.54
4	-1.469	20.207	-2.555	0.056	0.005	1925.67	**-3841.36**	19.58
8	0.529	41.655	-3.210	0.040	0.006	1920.59	-3831.2	19.51
16	2.628	60.088	-3.588	0.026	0.007	1913.37	-3816.74	19.45
IBM		QLB(30)	=1047.08				Real price	=107.06
1	-2.347	8.3960	-1.177	0.043	-0.001	2297.29	**-4584.60**	105.44
2	-1.547	29.958	-0.805	0.039	0.001	2294.93	-4579.87	105.36
4	0.316	62.380	-2.411	0.030	0.002	2288.43	-4566.86	105.41
8	1.590	82.893	-4.796	0.028	0.004	2285.42	-4560.85	105.24
16	1.242	89.824	-2.346	0.041	0.002	2281.99	-4553.99	105.41
DELL		QLB(30)	=1045.12				Real price	=27.18
1	-3.590	9.207	-3.385	0.089	0.006	1855.81	-3701.63	26.60
2	-1.133	34.127	-5.073	0.064	0.008	1865.98	-3721.97	26.68
4	-1.120	35.631	-4.981	0.066	0.008	1872.97	**-3735.95**	26.69
8	-1.736	39.809	-5.399	0.083	0.009	1870.08	-3730.17	26.68
16	0.137	55.953	-3.171	0.074	0.006	1871.40	-3732.81	26.66

Table 4.4: Parameter estimation under stationary GH-ARCH modelling. The MC prices were based on 20000 simulations. The forecasted period was set to 15 trading days. Models that minimize AIC information criteria are in bold. The overall time, for estimation and simulation, never exceeded 5 sec. QLB stand for the Ljung-Box test with 30 lags for no autocorrelation in the squared returns. The 5% critical value is 43.77.

CHAPTER 5

DISTRIBUTION OF THE INNOVATION PART OF A SD RANDOM VARIABLE

The innovation random variable for a non-negative self-decomposable random variable is fundamental for the analysis of models with SD invariant distributions such as OU-type processes. Section 5.1 gives the necessary tools for depicting such innovation term. This innovation random variable can have a compound Poisson distribution. In this case, Section 5.2 provides with the distribution for the compounded variable. Section 5.3 gives a latent representation for this random variable that is useful for simulation. When the innovation random variable does not have a compound Poisson representation, Section 5.4 presents a compound Poisson approximation for which the density function of the compounded variable is also available. These results can be used in the simulation of Ornstein-Uhlenbeck type processes with given marginal distributions. Previously, simulation of such processes has used the inverse of the corresponding tail Lévy measure. In Section 5.5 we show that this approach corresponds to the use of an inverse CDF method of a certain distribution. With knowledge of this distribution and hence density function, the sampling procedure is open to direct sampling methods. [1]

5.1 INTRODUCTION

Recent interest has focused on Ornstein-Uhlenbeck type processes and their application in stochastic volatility models. This application relies mainly in the subordination of Brownian motion with self-decomposable (SD) processes as operational time. See

[1] The material presented in this chapter will be appearing in Mena and Walker [74].

Sato [92] for more on subordination using SD processes. A detailed treatment of an application in stochastic volatility models is found in Barndorff-Nielsen and Shephard [13]. An important issue for volatility models is their simulation. OU-type processes are stationary processes driven by a positive Lévy process, without Gaussian components, and with SD marginals, see Section 2.6. Therefore, suitable representations of such processes (similarly SD random variables) can lead to suitable simulation techniques. For a detailed exposition of general Lévy process representations, see Rosiński [86]. A common problem for the techniques presented in [86] is the need for the inversion formula of the corresponding tail Lévy measure. The material on this section leads to alternative approaches.

Here, we discuss the simulation of the innovation random variable associated with SD random variables having tail Lévy measure of the form

$$N((x,\infty)) = \int_x^\infty \frac{1}{y} G(y)\, dy,$$

for some non-increasing function G. When $G(0)$ is finite, the innovation random variable is compound Poisson and we find the density function of the compounded variable. When $G(0) = \infty$ we use novel approximation methods based on the findings in the finite case in order to approximately sample the innovation variable. Our approach relies on the direct sampling of density functions rather than the use of inverse techniques.

A non-negative random variable X is said to be self-decomposable if for all $0 < \rho < 1$ there exists a (innovation) random variable W_ρ such that

$$X \stackrel{\mathrm{d}}{=} \rho X + W_\rho,$$

Where $\stackrel{\mathrm{d}}{=}$ denotes equal in distribution. Such a random variable X is infinitely divisible (see, for example, Vervaat [99]) and if $\inf\{x : \mathrm{P}(X \leq x) > 0\} = 0$ (assumed without loss of generality) then X has log-Laplace transform given by

$$\begin{aligned} -\log \mathbb{E}\left[e^{-\theta X}\right] &= \int_0^\infty \left(1 - e^{-\theta u}\right) dN(u) \\ &= \theta \int_0^\infty e^{-\theta y}\, \overline{N}(y)\, dy, \end{aligned}$$

where $N(\cdot)$ is a (Lévy) measure on $(0,\infty)$, $\overline{N}(y) = N((y,\infty))$, satisfying $N(1,\infty) < \infty$

and $\int_0^1 u\, dN(u) < \infty$, or in terms of the tail Lévy measure $\int_0^1 \overline{N}(u)\, du < \infty$. If we denote by $\mathcal{L}_Z(\theta)$ the Laplace transform of the random variable Z, then

$$\begin{aligned} \mathcal{L}_{W_\rho}(\theta) &= \frac{\mathcal{L}_X(\theta)}{\mathcal{L}_X(\rho\theta)} \\ &= \exp\left\{-\theta \int_0^\infty (e^{-\theta y} - \rho e^{-\theta\rho y})\overline{N}(y)\, dy\right\} \\ &= \exp\left\{-\theta \int_0^\infty e^{-\theta y}\overline{N}_{W_\rho}(y)\, dy\right\}, \end{aligned} \tag{5.1}$$

where $\overline{N}_{W_\rho}(y) = N((y, y/\rho)) = \overline{N}(y) - \overline{N}(y/\rho)$, is the Lévy measure corresponding to W_ρ and satisfies the same conditions as N.

In the case when X is a non-negative SD random variable, the tail Lévy measure takes the form

$$\begin{aligned} \overline{N}(x) &= \int_x^\infty \frac{1}{y} G(y)\, dy, \\ \overline{N}(x) &= \int_x^{\sim} \frac{1}{y} G(y)\, dy, \end{aligned}$$

for $x > 0$ and G is a decreasing function. See Sato [91], Section 53.

If $G(0) := \lim_{y \downarrow 0} G(y) < \infty$, we can set $S(y) := G(y)/G(0)$ as a well-defined survival function corresponding to a distribution function $F(y)$. With this notation, $\overline{N}(x)$ can be written as

$$\overline{N}(x) = \tau \int_x^\infty \frac{1}{y} S(y)\, dy, \tag{5.2}$$

where $\tau = G(0)$.

We will also consider, in Section 5.4, the case when $G(0) = \infty$. See Bondesson [18] for background and particular examples relating to these type of Lévy measures. A representation of X as a shot-noise random variable with exponential response is available;

$$X = \sum_{i=1}^{\infty} V_i\, e^{-T_i},$$

where $\{T_i\}$ denotes the sequence of points of a stationary Poisson process with intensity 1 and the V_i are independent and identically distributed from F. See Vervaat [99] for more details. On the other hand, a representation of X due to Ferguson and Klass [37],

is given by

$$X = \sum_{i=1}^{\infty} J_i,$$

where $N(J_i, \infty) = T_i$. Clearly, the latter approach needs the inverse of the tail Lévy measure, which for more relevant cases is not analytically available.

A problem considered by a number of authors in the early eighties concerned representations of the innovation random variable W_ρ. Suitable representations may lead to alternative simulation methods. Lawrance [58] found a representation of W_ρ when X is gamma distributed, say $\text{Ga}(\tau, 1)$, in terms of a compound Poisson distribution,

$$W_\rho = \sum_{i=1}^{k} Y_i \quad \text{and} \quad k \sim \text{Po}(-\tau \log \rho),$$

where the Y_i are independent and identically distributed random variables. Lawrance [58] provided the Laplace transform for Y, given by

$$\mathbb{E}\left[e^{-\theta Y}\right] = 1 - \frac{\log\left(\frac{1+\rho\theta}{1+\theta}\right)}{\log \rho},$$

and the result that $Y \stackrel{\text{d}}{=} \rho^U E$, where U is a uniform random variable from $[0, 1]$ and E an exponential random variable with mean 1, independent of U. This clearly gives an easy way to simulate random variates from W_ρ.

The main objective of this chapter is to generalize the Lawrance [58] result to a wider family of positive self-decomposable distributions. The innovation random variables considered in the next section are also compound Poisson and we provide the density function for the compounded variable Y explicitly.

5.2 FINITE ACTIVITY CASE $G(0) < \infty$

Here we focus on the particular case of self-decomposable random variables, for which the Lévy measure may be expressed as (5.2) with $G(0) < \infty$. We will deal with the infinite case in Section 5.4.

Theorem 5.1. The distribution for the innovation variable W_ρ of a self-decomposable random variable X, with Lévy measure expressed as in (5.2), can be represented as a compound Poisson random variable

$$W_\rho = \sum_{i=1}^{k} Y_i,$$

where $k \sim \text{Po}(-\tau \log \rho)$. Furthermore, the compounded variable Y has density function

$$h(y) = \frac{1}{y \log \rho} \{S(y/\rho) - S(y)\},$$

where $S(y) = G(y)/G(0)$, and h has distribution function

$$H(v) = 1 + \frac{1}{\log \rho} \int_v^{v/\rho} \frac{S(y)}{y}\, dy.$$

PROOF. First note that from assumption (5.2) we have

$$\overline{N}_{W_\rho}(v) = N((v, v/\rho)) = \tau \int_v^{v/\rho} \frac{1}{y} S(y)\, dy. \tag{5.3}$$

Hence the total mass for the measure N_{W_ρ} is given by

$$\begin{aligned} N_{W_\rho}((0,\infty)) &= \lim_{v \downarrow 0} \overline{N}_{W_\rho}(v) \\ &= \tau \lim_{v \to 0} \left\{ S(y) \log(y) \,|_v^{v/\rho} - \int_v^{v/\rho} \log(y)\, dS(y) \right\} \\ &= -\tau \log \rho. \end{aligned}$$

Therefore, normalizing $\overline{N}_W$, we can define

$$\overline{H}(v) := \frac{\overline{N}_{W_\rho}(v)}{-\tau \log \rho} = \frac{\int_v^{v/\rho} \frac{1}{y} S(y)\, dy}{-\log \rho}$$

and

$$H(v) = 1 + \frac{1}{\log \rho} \int_v^{v/\rho} \frac{1}{y} S(y)\, dy.$$

The density function, corresponding to $H(\cdot)$, is given by

$$h(y) = \frac{1}{y \log \rho} \{S(y/\rho) - S(y)\}.$$

We now need to show that $H(\cdot)$ is a well-defined distribution function on $(0, \infty)$.

1. $h(\cdot)$ is nonnegative
2. $\overline{H}(v)$ is non-increasing since $\overline{N}'_{W_\rho}(v) = \tau v^{-1} \{S(v/\rho) - S(v)\} < 0$
3. If $S(\cdot)$ is continuous at 0 then $H(0) = 0$
4. It remains to prove that $\overline{H}(y) \to 0$ as $y \to \infty$, that is $\overline{N}_{W_\rho}(v) \to 0$ as $v \to \infty$. For any $\epsilon > 0$ there exists $v_\epsilon > 0$, with such that $G(v) < \epsilon$ for all $v > v_\epsilon$. If $\epsilon' > 0$ and v_ϵ as above with $\epsilon = -\epsilon' \log \rho$, then

$$\int_v^{v/\rho} \frac{G(x)}{x}\, dx < \int_v^{v/\rho} \frac{\epsilon'}{x}\, dx = \epsilon.$$

Once shown that $H(\cdot)$ is a well-defined distribution function we can verify that W_ρ is distributed as a compound Poisson random variable with compounded variable $Y \sim H(\cdot)$. Now

$$\mathbb{E}\left[e^{-\theta W_\rho}\right] = \mathbb{E}\left[[\mathcal{L}_Y(\theta)]^k\right] = \exp\left\{-\lambda\left(1 - \int_0^\infty e^{-\theta y} h(y)\, dy\right)\right\},$$

where $\lambda = -\tau \log \rho$, and

$$\begin{aligned}\int_0^\infty e^{-\theta y} h(y)\, dy &= \theta \int_0^\infty e^{-\theta y} H(y)\, dy \\ &= 1 - \frac{\theta}{\lambda} \int_0^\infty e^{-\theta y} \overline{N}_{W_\rho}(y)\, dy.\end{aligned}$$

Therefore,

$$\mathrm{E}\left[e^{-\theta W_\rho}\right] = \exp\left\{-\theta \int_0^\infty e^{-\theta y} \overline{N}_{W_\rho}(y)\, dy\right\}. \tag{5.4}$$

Expression (5.4) coincides with (5.1), then $H(u)$ is as stated. This completes the proof. □

If $\mathrm{P}(Y = 0) = 0$ then $\mathrm{P}(W_\rho = 0) = \mathrm{P}(k = 0) = e^{-\lambda} = \rho^\tau$. If $G(0) = \infty$ the representation presented in Theorem 5.1 is not valid, since $S(\cdot)$ is not a survival function. An example of this case is given when we assume that X is inverse Gaussian (IG). See Example 5.3 below, for more on this.

Example 5.1. (Lawrance [58]) If X is $\mathrm{Ga}(\alpha, 1)$ then $N(x, \infty) = \alpha \int_x^\infty y^{-1} e^{-y}\, dy$ and so

$$h(y) = \frac{1}{-y \log \rho} \left\{ e^{-y} - e^{-y/\rho} \right\}.$$

This has the Laplace transform given earlier in Section 5.1; that is,

$$\int_0^\infty e^{-\theta y} h(y)\, dy = 1 - \frac{\log\left(\frac{1+\rho\theta}{1+\theta}\right)}{\log \rho}.$$

This density generalizes the exponential density which arises as $\rho \to 1$, which corresponds to the case where the "thinned" part ρX tends to X and then the innovation term tends to zero. All the moments exist and $\mathbb{E}[Y^r] = -(r-1)!(1-\rho^r)/\log \rho$. ○

Example 5.2. Now let us take $G(y) = e^{-y^\xi}$, $\xi > 0$. Note that for $\xi = 1$ we are in the case of Example 5.1. The function $S(x) := G(x)$ is the survival function corresponding to a random variable $V \sim \mathrm{Weibull}(\xi, 1)$. Clearly, for this case, we can verify

$$\int_0^1 un(u)\, du = \tau \int_0^1 e^{-u^\xi}\, du < \infty \text{ and } N(1, \infty) = \tau \int_1^\infty \frac{e^{-u^\xi}}{u}\, du < \infty,$$

leading to a valid infinitely divisible random variable X. Here the compounded random variable in the representation for W_ρ has density function

$$h(y) = \frac{1}{-y \log \rho} \left\{ e^{-y^\xi} - e^{-(y/\rho)^\xi} \right\}.$$

○

5.3 SAMPLING Y

Let us assume that $S(\cdot)$ has a density function with respect to Lebesgue measure. Then we can write the density function $h(y)$, of the compounded random variable Y, as

$$h(y) = \frac{1}{-y \log \rho} \int_y^{y/\rho} f(\xi)\, d\xi,$$

where $f(\cdot)$ denotes the density function corresponding to $S(\cdot)$. An augmentation of this random variable can be done by the following change of variable $\xi = y/z$, we get

$$h(y) = \frac{1}{-\log \rho} \int \frac{f(y/z)\, \mathbb{1}(\rho < z < 1)}{z^2}\, dz.$$

Thus let us consider the joint density function given by

$$h(y, z) = \frac{f(y/z)\, \mathbb{1}(\rho < z < 1)}{-z^2 \log \rho},$$

where the corresponding marginal density function for the latent variable Z is given by

$$h(z) = \frac{\mathbb{1}(\rho < z < 1)}{-z \log \rho}.$$

A random variable Z from this density function can be taken as $Z = \rho^U$, where U is uniform from $[0, 1]$. Consequently, we can deduce that

$$Y \stackrel{\mathrm{d}}{=} V \rho^U, \tag{5.5}$$

where $V \sim f$ and is independent of U. Representation (5.5) provides us with an easy way to simulate random variates Y, and therefore also W_ρ, Figure 5.1 illustrates some simulations of Y corresponding to Example 5.1 and Example 5.2.

5.4 INFINITE ACTIVITY CASE $G(0) = \infty$

Our aim here is to approximate the distribution of the underlying innovation random variable when W_ρ is not compound Poisson; when $G(0) = \infty$. It is well known that an approximation can be made via compound Poisson random variables. In Bondesson [18] this approximation was mentioned for general infinitely divisible Lévy processes. However, in the particular case of SD distributions, a different approximation is found to be useful.

Our method is now introduced. In the case when $G(0) = \infty$ we can approximate (5.3) with

$$\overline{N}^{\epsilon}_{W_\rho}(\nu) = \int_{\nu}^{\nu/\rho} \frac{G_\epsilon(y)}{y}\, dy,$$

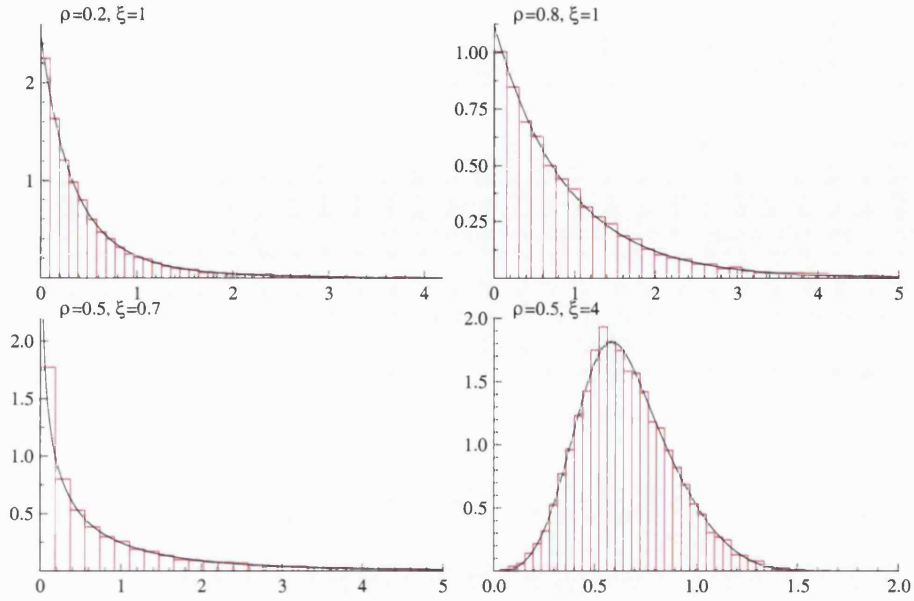


Figure 5.1: Simulations from the compounded random variable Y, using representation (5.5) with $V \sim \text{Weibull}(\xi, 1)$. $S(y) = e^{-y^{\xi}}$, 10,000 simulations. The solid lines represent the true densities.

where

$$G_\epsilon(y) = G(\epsilon)\, \mathbb{1}(y \leq \epsilon) + G(y)\, \mathbb{1}(y > \epsilon),$$

for $\epsilon > 0$. Here $G_\epsilon(0) = G(\epsilon) < \infty$ and $\overline{N}^{\epsilon}_{W_\rho}(\nu) \to \overline{N}_{W_\rho}(\nu)$ as $\epsilon \to 0$. See Figure 5.2. We could equally use the approximation $G^*_\epsilon(y) = G(y + \epsilon)$ but in this chapter we use $G_\epsilon(\cdot)$.

Proposition 5.1. When $G(0) = \infty$ the approximation W^ϵ_ρ for the innovation random variable converges weakly to W_ρ, that is,

$$W^\epsilon_\rho \xrightarrow{\text{d}} W_\rho, \text{ as } \epsilon \to 0.$$

Here $\xrightarrow{\text{d}}$ denotes convergence in distribution.

Proof. In order to proof convergence in distribution we use continuity theorem. Therefore, it is sufficient to show $\mathcal{L}_{W^\epsilon_\rho}(\theta) \to \mathcal{L}_{W_\rho}(\theta)$ as $\epsilon \to 0$. Using expression (5.1), we see

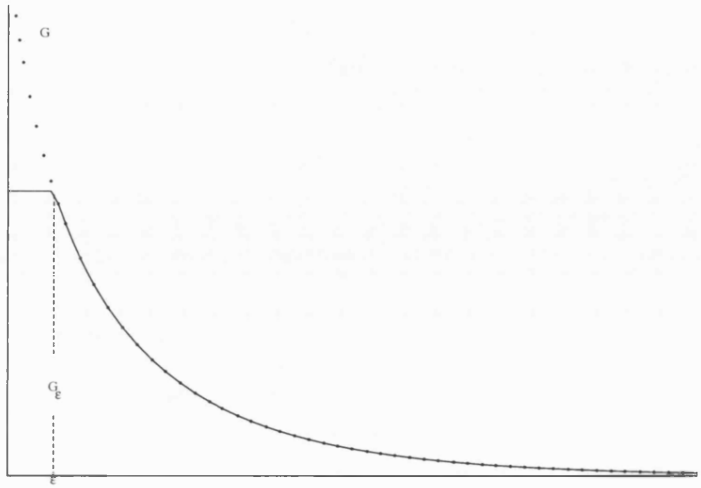


Figure 5.2: Approximation for the function G.

that convergence of Laplace transforms is the same as proving the following:

$$\lim_{\epsilon \to 0} \int_0^\infty e^{-\theta\nu}\, \overline{N}^{\epsilon}_{W_\rho}(\nu)\, d\nu = \int_0^\infty e^{-\theta\nu}\, \overline{N}_{W_\rho}(\nu)\, d\nu.$$

Notice that $\epsilon \mapsto \overline{N}^{\epsilon}_{W_\rho}(\cdot)$ is a decreasing function with $\overline{N}_{W_\rho}(\cdot)$ as the limit as $\epsilon \to 0$ and therefore the monotone convergence theorem applies and the result follows. □

For the approximated Lévy measure we have

$$N^{\epsilon}_{W_\rho}((0,\infty)) = \lim_{\nu \downarrow 0} \overline{N}^{\epsilon}_{W_\rho}(\nu) = -\tau_\epsilon \log \rho,$$

where $\tau_\epsilon = G_\epsilon(0) = G(\epsilon)$. Let us define $S_\epsilon(y) := G_\epsilon(y)/G(\epsilon)$, which is a survival function on $(0,\infty)$. Consequently, define

$$h^\epsilon(y) = \frac{S_\epsilon(y/\rho) - S_\epsilon(y)}{y \log \rho}, \tag{5.6}$$

and

$$W^\epsilon_\rho = \sum_{i=1}^{k_\epsilon} Y^\epsilon_i, \tag{5.7}$$

where $k_\epsilon \sim \mathrm{Po}(-\tau_\epsilon \log \rho)$ with the random variables Y^ϵ_i having density function given by (5.6). Clearly, W^ϵ_ρ has a compound Poisson distribution. The same arguments used in Theorem 5.1 follow in this case, leading to a well-defined probability density (5.6)

for the compounded random variable (5.7). As before, we can write $Y^\epsilon \stackrel{\mathrm{d}}{=} V_\epsilon \rho^U$, where V_ϵ has distribution function $F_\epsilon(x) = 1 - S_\epsilon(x)$.

The following proposition is useful for accuracy assessment of the approximation to the function $G_\epsilon(\cdot)$.

Proposition 5.2. The Lévy density $\eta^\epsilon(\nu)$ for the random variable $Z^\epsilon := W_\rho - W_\rho^\epsilon$ is given by

$$\eta^\epsilon(\nu) = \frac{G(\nu) - G(\nu/\rho)}{\nu} \mathbb{1}(0 < \nu < \epsilon\rho) + \frac{G(\nu) - G(\epsilon)}{\nu} \mathbb{1}(\epsilon\rho < \nu < \epsilon) \tag{5.8}$$

and therefore the expectation and variance are given by

$$\mathrm{E}[Z^\epsilon] = (1-\rho)\left\{\int_0^\epsilon G(y)dy - \epsilon G(\epsilon)\right\} \quad \text{and} \tag{5.9}$$

$$\mathrm{Var}[Z^\epsilon] = (1-\rho^2)\left\{\int_0^\epsilon yG(y)dy - \frac{\epsilon^2}{2}G(\epsilon)\right\}. \tag{5.10}$$

Proof. First notice that for $A \in \mathcal{B}(R^+)$ we have

$$\begin{aligned} N_W^\epsilon(A) &= \int_A \frac{G_\epsilon(y) - G_\epsilon(y/\rho)}{y} dy \\ &= \int_{A\cap(\epsilon,\infty)} \frac{G(y) - G(y/\rho)}{y} dy + \int_{A\cap[\epsilon\rho,\epsilon]} \frac{G(\epsilon) - G(y/\rho)}{y} dy. \end{aligned} \tag{5.11}$$

The first integral in (5.11) corresponds to the $\epsilon-$truncation of the Lévy measure. Therefore, the second integral represents the extra contribution of our approximation method. Hence the Lévy measure for the error is given by

$$\begin{aligned} N_W(A) - N_W^\epsilon(A) &= \int_{A\cap(0,\epsilon]} \frac{G(y) - G(y/\rho)}{y} dy - \int_{A\cap[\epsilon\rho,\epsilon]} \frac{G(\epsilon) - G(y/\rho)}{y} dy \\ &= \int_{A\cap(0,\epsilon\rho)} \frac{G(y) - G(y/\rho)}{y} dy - \int_{A\cap[\epsilon\rho,\epsilon]} \frac{G(y) - G(\epsilon)}{y} dy \end{aligned}$$

and thus the corresponding Lévy density (5.8) follows.

For a given $G(\cdot)$ we can compute the expectation and variance of the difference $W_\rho - W_\rho^\epsilon$.

$$\begin{aligned}
\mathrm{E}\left[W_\rho - W_\rho^\epsilon\right] &= \int_0^\infty y\,\eta^\epsilon(y)\,dy \\
&= \int_0^\epsilon G(y)\,dy - \int_0^{\epsilon\rho} G(y/\rho)\,dy - (1-\rho)\epsilon G(\epsilon) \\
&= (1-\rho)\left\{\int_0^\epsilon G(y)\,dy - \epsilon G(\epsilon)\right\}
\end{aligned}$$

and

$$\begin{aligned}
\mathrm{Var}\left[W_\rho - W_\rho^\epsilon\right] &= \int_0^\infty y^2\,\eta^\epsilon(y)\,dy \\
&= \int_0^\epsilon y\,G(y)\,dy - \int_0^{\epsilon\rho} y\,G(y/\rho)\,dy - (1-\rho^2)\,\frac{\epsilon^2}{2}\,G(\epsilon) \\
&= (1-\rho^2)\left\{\int_0^\epsilon yG(y)\,dy - \frac{\epsilon^2}{2}\,G(\epsilon)\right\}.
\end{aligned}$$

Note that the integrals in (5.9) and (5.10) are finite since the conditions on the Lévy measure $N(1,\infty) < \infty$ and $\int_0^1 u\,dN(u) < \infty$ are satisfied. Therefore it is always possible bound the integrals in (5.9) and (5.10). This give us a way two achieve, in mean, a desire accuracy in the approximation W_ρ^ϵ for W_ρ. Clearly both the mean and variance tend to zero as ϵ tend to zero. □

Example 5.3. If $X \sim \mathrm{IG}(\delta,\gamma)$ then the corresponding Lévy measure has density $n(x) = G(x)/x$, with

$$G(x) = \frac{\delta}{\sqrt{2\pi x}}\exp\left\{-\frac{\gamma^2 x}{2}\right\},$$

and clearly $G(0) = \infty$. In this case,

$$S_\epsilon(y) = \frac{\sqrt{\epsilon}\exp\left\{-\frac{\gamma^2}{2}(y-\epsilon)\right\}}{\sqrt{y}}\,\mathbb{1}(y > \epsilon) + \mathbb{1}(y \le \epsilon),$$

defines a survival function with corresponding density function

$$f_\epsilon(y) = \frac{\sqrt{\epsilon}\exp\left\{-\frac{\gamma^2}{2}(y-\epsilon)\right\}(1+\gamma^2 y)}{2y^{3/2}}\,\mathbb{1}(y > \epsilon). \tag{5.12}$$

The density (5.12) can be written as

$$f_\epsilon(x) \propto h(x)\, k(x)\, \mathbb{1}(x > 0), \qquad x = y - \epsilon \tag{5.13}$$

and

$$h(x) = \frac{\sqrt{\epsilon}}{2(x+\epsilon)^{3/2}}, \quad k(x) = e^{-\gamma^2 x/2}\left\{1 + \gamma^2(x+\epsilon)\right\}. \tag{5.14}$$

○

Hence, in order to simulate from the random variable V_ϵ with density (5.12) we can simulate from (5.13) and add ϵ. The decomposition in (5.13) allows us to use the acceptance-rejection method by simulating from $h(\cdot)$ in (5.14) and with acceptance criteria $U \leq k(x)/M$, $M = \sup_x k(x) = k(m)$ with $m = \max\{\gamma^{-2} - \epsilon, 0\}$ and U is an uniform $[0,1]$ random variate. See Rubinstein [87]. Therefore, to simulate from the random variable (5.7) we follow the next steps:

- For any $\epsilon > 0$ simulate $k_\epsilon \sim \text{Po}(-G(\epsilon)\log\rho)$.
- Simulate k_ϵ independent random numbers from an uniform distribution in $[0,1]$ and k_ϵ independent random numbers from V_ϵ as described above.
- Compute (5.7).

In order to illustrate this method graphically, notice that we can approximate a SD random variate X by simulating from $X^\epsilon = \rho X + W_\rho^\epsilon$. If $X \sim \text{IG}(\delta, \gamma)$ then $\rho X \sim \text{IG}(\delta\sqrt{\rho}, \gamma/\sqrt{\rho})$, therefore, a random variate from X can be approximated as the sum of a random number from $\rho X \sim \text{IG}(\delta\sqrt{\rho}, \gamma/\sqrt{\rho})$ and a random number from W_ρ^ϵ.

As an alternative to our direct sampling method, V_ϵ can be simulated using the inverse CDF method. This corresponds to the standard inverse tail Lévy measure method. Figure 5.3 reports some simulations using the inverse CDF method and our method. The reduction in computer-time is considerable; the simulation of 3,000 random variables took 0.36 seconds for the inverse CDF method compared with 0.06 seconds for ours. All computations were done in Ox; see Doornik [28].

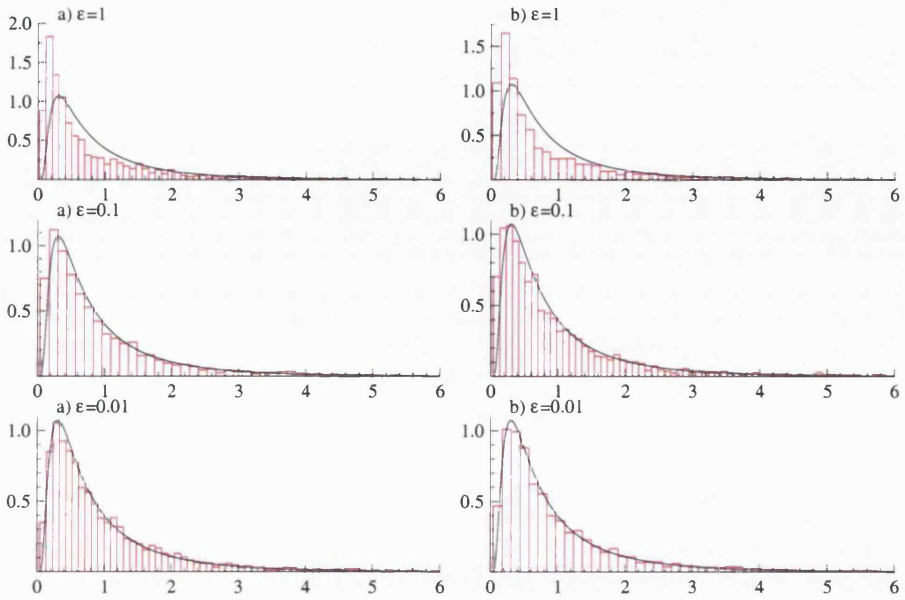


Figure 5.3: Approximation of self-decomposable random variable IG(1,1) using 3000 simulations of W_ρ^ϵ and ρX with $\rho = 0.5$. The simulations for the compounded random variable were done using the inverse CDF method (a) and the acceptance-rejection method (b). The solid line represents the true density.

5.5 RELATION WITH OU-TYPE PROCESSES AND THEIR REPRESENTATIONS

In Barndorff-Nielsen and Shephard [13] a crucial feature, e. g. to predict future levels of volatility, is the simulation of the innovations for OU-type processes. OU-type processes are stationary processes with marginal distributions given by SD distributions (see, for example, Sato [91]). Here we focus on OU-type processes, when a prior choice of the stationary distribution is given. We have seen, in Section 2.6, that an OU-type process can be represented as follows

$$X(t) \;=\; \rho^t X(0) + \rho^t \int_0^{at} e^s \, dL(s) \quad \text{for } \rho = e^{-a}, \quad a > 0, \tag{5.15}$$

where $L(\cdot)$ is a Lévy process on $(0, \infty)$. See Wolfe [107]. In order to simulate from the innovation part (the second term in (5.15)), Barndorff-Nielsen and Shephard [13] made use of the following result;

$$\int_0^{\Upsilon} f(s)\, dL(s) \stackrel{\mathrm{d}}{=} \sum_{i=1}^{\infty} M^{-1}(\xi_i/\Upsilon)\, f(\Upsilon r_i),$$

where M^{-1} denotes the inverse of the tail Lévy measure corresponding to $L(1)$. Here $\{\xi_i\}$ and $\{r_i\}$ are two independent sequences of random variables, with the r_i being independent and identically distributed from the uniform distribution on $[0,1]$ and $\xi_1 < \cdots < \xi_i < \cdots$ are the jump times of a Poisson process with intensity 1. It is worth noting that the above result can be seen as particular case of a representation given by Ferguson and Klass [37] (see Walker, [101]). For the OU-innovation this representation simplifies as

$$\begin{aligned} X(t) &= \rho^t X(0) + \rho^t \int_0^{at} e^s\, dL(s) \\ &\stackrel{\mathrm{d}}{=} \rho^t X(0) + \sum_{i=1}^{\infty} \rho^{(1-r_i)t}\, M^{-1}(\xi_i/at) \\ &\stackrel{\mathrm{d}}{=} \rho^t X(0) + \sum_{i=1}^{\infty} \rho^{tr_i}\, M^{-1}(\xi_i/at). \end{aligned} \tag{5.16}$$

Example 5.4. Consider a Lévy process L with gamma $\mathrm{Ga}(\tau, 1)$ increments, thus

$$M^{-1}(x) = \max\{0, -\log(x/\tau)\}.$$

In this case the innovation part (the second summand in (5.16)) is represented as

$$\begin{aligned} \rho^t \int_0^{at} e^s\, dL(s) &\stackrel{\mathrm{d}}{=} \sum_{i=1}^{\infty} \rho^{tr_i} \log(1/c_i)\, \mathbb{I}(0 < c_i < 1) \\ &= \sum_{i=1}^{N(1)} \rho^{tr_i} \log(1/c_i), \end{aligned}$$

where $c_1 < \cdots < c_i < \cdots$ are the jump times of a Poisson process with intensity $a\tau t$ (or $-t\tau \log(\rho)$) and $N(1)$ corresponding number of jumps before 1. Let k be the number of jumps before 1, then given k the $\{c_i\}$ are independent and identically distributed from a uniform distribution in $[0,1]$. Hence if we define $V_i = -\log(c_i)$ thus $V_i \sim \mathrm{Exp}(1)$ and therefore

$$\rho^t \int_0^{at} e^s \, dL(s) \stackrel{\mathrm{d}}{=} \sum_{i=1}^{N(1)} \rho^{tr_i} V_i,$$

which, for $t = 1$, is exactly the representation provided in Example 5.1. ∘

In Example 5.4 we are in the case where $G(0) < \infty$ so the innovation variable has a compound Poisson distribution. For this case neither the method used by Barndorff-Nielsen and Shephard [13] nor our approach needs a truncation. However, in cases where $G(0) = \infty$ the method used in [13] needs the truncation of the summation in the series representations in addition to the inverse of the tail Lévy measure. In our approach the computation of the latter function can be circumvented by switching from the inverse CDF method to a different one such as the acceptance-rejection. The truncation in our method translates to chose a suitable value for ϵ, for which Proposition 5.2 can help to asses suitable errors.

Knowledge of the distribution for the innovation random variable (or an approximation to it) of a self-decomposable random variable gives new ways of simulating the innovation part of an OU-type process. The point of view presented in this chapter allows us to use any suitable random variable simulation method (not only inverse methods). Figure 5.4 shows some simulations of an OU-type processes with stationary distribution being IG(1, 1). The simulations were implemented using the approximation scheme described in Section 5.4 and Example 5.3. The corresponding autocorrelation functions are shown in Figure 5.5, notice how a better approximation to the process leads to higher autocorrelations.

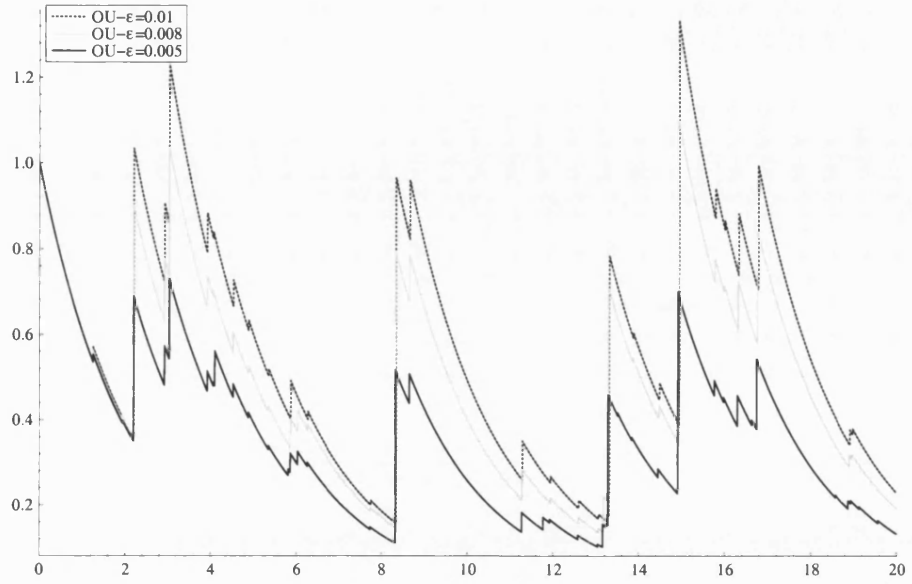


Figure 5.4: Simulations of OU-type processes with stationary distribution IG(1, 1) and $\rho = 0.6$. The simulations were done on a time-grid of 0.001 and for different values of ε. All simulations were started at $X(0) = 1$.

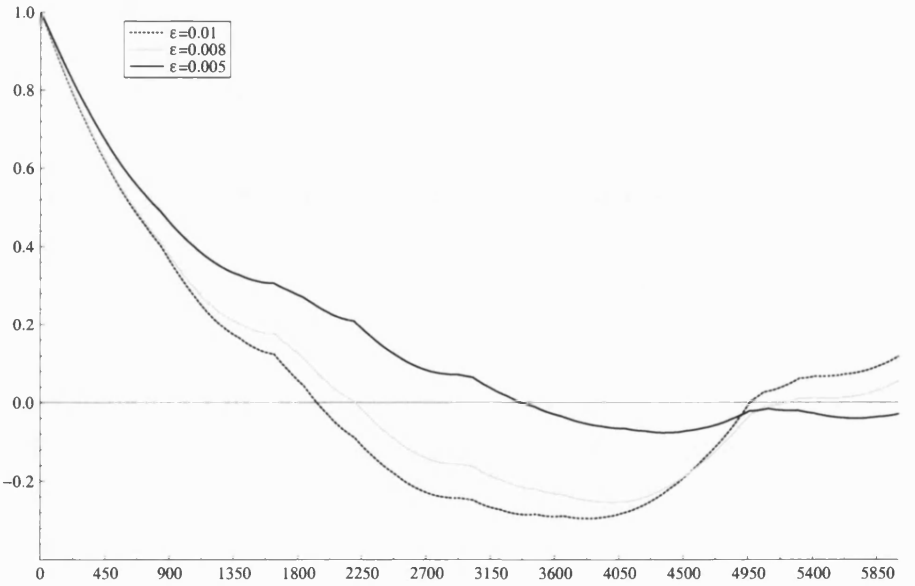


Figure 5.5: ACF's for the simulations in Figure 5.4.

CHAPTER 6

LATENT STRUCTURE BASED MODELS: CONTINUOUS TIME

In this chapter, we consider the continuous-time Gibbs sampler type construction introduced in Chapter 3. Main interest is devoted to the weak representation of diffusion processes. Section 6.1 discuses the main complexity when considering continuous-time models. In Section 6.2 a latent representation of Gaussian OU processes is presented. We have seen in Chapters 3 and 4 that the specification of the parametric family $F_{Y|X}$ is essential for the underlying construction. In Sections 6.3 and 6.5 we fix this parametric family to be Poisson and explore the resulting models. The final part of this Chapter, Section 6.6 provides with an estimation example.

6.1 INTRODUCTION

In Example 3.3 we have seen that using the Gibbs type construction we can construct continuous-time Markov models. In particular, Example 3.3 provides with an instance of a continuous-time Markov chain. In this chapter, we concentrate our efforts to the construction of stationary Markov models where the invariant measure is diffuse, particularly we are interested in the identification of some diffusion models. Hence, using the *general setting with target process* $X = \{X(t); t \in \mathbb{R}_+\}$, described in Chapter 3, we are given two sets of conditional probabilities (3.20) and (3.21). Then, the objective is to construct suitable continuous-time homogeneous Markov transition probabilities via a latent structure, such that the associated process has a given invariant measure

Q_X. Following Chapter 3, this can be done by using the transition probability

$$\mathrm{P}_t(x, B) := \int_S F_{X_t|Y_t=y}(B)\, F_{Y_t|X_0=x}(dy), \quad x \in \mathrm{E}, B \in \mathcal{E}, \tag{6.1}$$

where $F_{X_t|Y_t}$ and $F_{Y_t|X_0}$ arise from a Gibbs sampler type update given by (3.16).

In the discrete-time setting the one-step transition probabilities are constructed such that the Chapman-Kolmogorov equations (2.5) are immediately satisfied. The difference in the continuous-time setting is that we should incorporate the time information in the conditional distributions $F_{X_t|Y_t}$ and $F_{Y_t|X_0}$ in such a way that (6.1) satisfies the Chapman-Kolmogorov equations. Finding suitable ways of incorporating time-information in the underlying conditional distributions is not trivial. One possibility may be to obtain the n-step transition probability and from it to construct a continuous-time transition probability. Namely, given a valid n-step time-homogenous transition probability $P^n(x, A)$, $n \in \mathbb{N}$, with a tractable expression, it may be easier to get an insight for the transition probability on a more general time index, in general $\mathbb{R}_+$. This works for some simple cases as the one presented in Section 6.2. However, for cases where the one-step transition probabilities do not have a simple form, the integration required to get the n-step transition probabilities - see equation (4.2) - is generally hard to accomplish analytically. Alternatively, one may allow the underlying conditional probabilities to depend on a certain time-function, and then find the proper forms for it such that Chapman-Kolmogorov equations are satisfied. More precisely, once the parametric families for the conditionals are fixed, we allow one (or more) of their parameters to depend on time through an unknown function ϕ_t, for instance we might have $F^{\phi_t}_{Y_t|X_0}$ and $F^{\phi_t}_{X_t|Y_t}$, and therefore find a suitable form for ϕ_t such that (6.1) satisfies the Chapman-Kolmogorov equations.

Once a suitable transition corresponding to the target process has been constructed, we can relate it to a diffusion process (when possible) by finding the infinitesimal mean and variance coefficients; see Chapter 2, equations (2.18) and (2.19). In fact, provided that a choice of the function ϕ_t leads to a model that satisfies the Dynkin condition (2.16), then a diffusion model can be associated. An important issue to emphasize is that the constructed transition only gives the behavior of the related diffusion in the interior of the state space E. The behavior at the boundary points must be analyzed separately. We refer to Karlin and Taylor [53] for classification of boundaries.

It is also worth to remember that we are focusing on the separable modification of a process, as explained in Chapter 2. In the discrete-time case the trajectories or paths are clearly not continuous. However, in the continuous-time case we can have different types of continuity. Introducing a process with the construction at issue, we cannot tell much about the path continuity. Therefore, in order to associate the constructed process with a process in continuous time, in particular a diffusion process, further assumptions on the path continuity have to be made. It suffices to assume that our constructed target process is standard.

6.2 GAUSSIAN ORNSTEIN-UHLENBECK MODEL

Let us assume that we want to represent a Gaussian OU diffusion using a Gibbs type construction. It is well known that OU processes are essentially the only stationary Gaussian processes with the Markov property, therefore all finite dimensional distributions of a OU process must be Gaussian, see Examples 2.1 and 2.2. This implies that all the conditionals (transitions) are also Gaussian.

In Example 4.1 we saw that a discrete-time model with invariant distribution $Q_X = \mathrm{N}(0,1)$ can be generated by assuming that $F_{Y|X=x} = \mathrm{N}(x, r)$, $r > 0$. We also saw that the n-step transition probability is given by

$$\mathrm{P}^n(x, \cdot) = \mathrm{N}\left(\frac{x}{(r+1)^n}, 1 - (r^2 + 2r + 1)^{-n}\right), \tag{6.2}$$

for $n \in \mathbb{N}_0$.

A natural question to ask in the continuous-time context is whether the transition (6.2) can have a continuous time counterpart. Making the appropriate changes in notation, we assume the following transition holds

$$\mathrm{P}_t(x, \cdot) = \mathrm{N}\left(\frac{x}{(r+1)^t}, 1 - (r^2 + 2r + 1)^{-t}\right). \tag{6.3}$$

for $t > 0$ and $x \in \mathrm{intE}$, where $\mathrm{E} = \mathbb{R}$. The objective then, is to verify if the above transition probability can be seen as the weak solution of an OU process, after a suitable re-parametrization.

Proposition 6.1. The transition probability defined by (6.3) satisfies the Chapman-Kolmogorov equations.

Proof. Let $C_X(\cdot)$ denote the characteristic function of the random variable X. Hence, the Chapman-Kolmogorov equations can be written as

$$C_{X_{t+s}|X_0=x}(\xi) = \mathbb{E}\left[C_{X_{t+s}|X_s}(\xi) \mid X_0 = x\right].$$

Put $\eta_t := 1 - (r+1)^{-2t}$, then

$$C_{X_t|X_0=x}(\xi) = \exp\left\{\frac{ix\xi}{(r+1)^t} - \frac{\xi^2\eta_t}{2}\right\}$$

and

$$\begin{aligned}
\mathbb{E}\left[C_{X_{t+s}|X_s}(\xi) \mid X_0 = x\right] &= \exp\left\{-\frac{\xi^2}{2}\eta_t\right\} C_{X_s|X_0=x}\left(\frac{\xi}{(r+1)^t}\right) \qquad (6.4)\\
&= \exp\left\{-\frac{\xi^2}{2}\left(\eta_t + \frac{\eta_s}{(r+1)^{2t}}\right)\right\} \exp\left\{\frac{ix\xi}{(r+1)^{t+s}}\right\}\\
&= \exp\left\{\frac{ix\xi}{(r+1)^{t+s}} - \frac{\xi^2}{2}\eta_{t+s}\right\}\\
&= C_{X_{t+s}|X_0=x}(\xi).
\end{aligned}$$

Equality (6.4) is valid due to the homogeneity of the transition probability. The uniqueness of characteristic functions leads to the stated result. □

In order to associate the above model with a diffusion process, we can get the infinitesimal drift and diffusion coefficients using (2.18) and (2.19) respectively, that is

$$\begin{aligned}
\mu(x) &= \lim_{h\downarrow 0}\frac{1}{h}\mathbb{E}[X_{t+h} - X_t \mid X_t = x]\\
\sigma^2(x) &= \lim_{h\downarrow 0}\frac{1}{h}\mathbb{E}[(X_{t+h} - X_t)^2 \mid X_t = x].
\end{aligned}$$

In this case, using the homogeneity of the constructed process, we have

$$\begin{aligned}
\mu(x) &= \lim_{h\downarrow 0}\frac{1}{h}\left\{\mathbb{E}[X_h \mid X_0 = x] - x\right\} = x\lim_{h\downarrow 0}\frac{1}{h}\{(r+1)^{-h} - 1\}\\
&= -x\ln(r+1) \quad \text{and} \qquad (6.5)
\end{aligned}$$

$$\begin{aligned}\sigma^2(x) &= \lim_{h\downarrow 0}\frac{1}{h}\left\{\mathbb{E}[X_h^2 \mid X_0 = x] - 2x\mathbb{E}[X_h \mid X_0 = x] + x^2\right\} \\ &= \lim_{h\downarrow 0}\frac{1}{h}\left\{1-(r+1)^{-2h} + x^2(r+1)^{-2h} - 2x^2(r+1)^{-h} + x^2\right\} \\ &= \lim_{h\downarrow 0}\frac{1}{h}\left\{1-(r+1)^{-2h}\right\} + x^2\lim_{h\downarrow 0}\frac{1}{h}\left\{(r+1)^{-2h} - 2(r+1)^{-h} + 1\right\} \\ &= 2\ln(r+1).\end{aligned}$$

To ensure that the constructed process can be seen as a diffusion process, the Dynkin condition must be satisfied. It is enough to corroborate, as seen in equation (2.17), that the absolute value of a central infinitesimal moment higher than two is zero. Here, to avoid the absolute value we use the fourth moment[1], that is

$$\lim_{h\downarrow 0}\frac{1}{h}\mathbb{E}[(X_{t+h} - X_t)^4 \mid X_t = x] = 0, \tag{6.6}$$

for any $x \in \mathbb{R}$. Expanding the expectation in (6.6) we get

$$\lim_{h\downarrow 0}\frac{1}{h}\left\{\mathbb{E}_x[X_h^4] - 4x\mathbb{E}[X_h^3] + 6x^2\mathbb{E}[X_h^2] - 4x^3\mathbb{E}[X_h] + x^4\right\},$$

where $\mathbb{E}_x[\cdot]$ denotes the expectation taken with respect to (6.3). The third and fourth moments of a $\mathrm{N}(\mu,\nu)$ are given by $\mu(3\nu+\mu^2)$ and $3\nu^2+6\nu\mu^2+\mu^4$ respectively. Therefore, replacing the latter quantities with the corresponding moments and after some algebra, the above limit is simplified as follows

$$\begin{aligned} & 3\lim_{h\downarrow 0}\frac{1}{h}\left\{1-2(r+1)^{-2h}+3(r+1)^{-4h}\right\} \\ + \; & 6x^2\lim_{h\downarrow 0}\frac{1}{h}\left\{(r+1)^{-2h} - 2(r+1)^{-h} + 2(r+1)^{h} - (r+1)^{2h}\right\} \\ + \; & x^4\lim_{h\downarrow 0}\frac{1}{h}\left\{(r+1)^{-4h} - 4(r+1)^{-3h} + 6(r+1)^{-2h} - 4(r+1)^{-h} + 1\right\} \\ = \; & 0. \end{aligned} \tag{6.7}$$

Hence the Dynkin condition is satisfied. Having showed this and assuming that the constructed process is standard, the target process X can be seen as a version corresponding to the weak solution of the following stochastic differential equation

$$dX_t = -\ln(r+1)X_t\,dt + \sqrt{2\ln(r+1)}\,dW_t, \tag{6.8}$$

[1]Although, the Dynkin condition might be checked with moments higher than three, the fourth moment it is the common choice.

where $r > 0$ and W_t denotes a one-dimensional Brownian motion.

If we let $r = e - 1$, then the transition probability (6.3) is given by $\mathrm{P}_t(x, \cdot) = \mathrm{N}\left(xe^{-t}, 1 - e^{-2t}\right)$, which is recognizable as the transition probability corresponding to the solution of

$$dX_t = -X_t\, dt + \sqrt{2}\, dW_t. \tag{6.9}$$

Equation (6.9) can be seen as a particular case of an Ornstein-Uhlenbeck process given as the solution to

$$dX_t = -\alpha X_t\, dt + \sigma\, dW_t \tag{6.10}$$

with $\alpha = 1$ and $\sigma = \sqrt{2}$.

On the same line as Proposition 6.1 and equations (6.5–6.7), we can get a more general version of Gaussian OU process. Let us assume that we want to construct a stationary Markov process $\{X_t, t \in \mathbb{R}\}$ with invariant distribution given by

$$Q_X(\cdot) = \mathrm{N}\left(\mu, \frac{\sigma^2}{2\alpha}\right), \tag{6.11}$$

where $\alpha, \sigma^2 > 0$ and $\mu \in \mathbb{R}$. Using the notation of Chapter 3, we have that $\mathrm{E} = \mathbb{R}$, $\mathrm{S} = \mathbb{R}$ and $\mathcal{T} = \mathbb{R}_+$. By a suitable re-parametrization of the choice for the conditional distribution $F_{Y|X}$ made in Examples 3.2 and 4.1, we assume that

$$F_{Y_t|X_0}(\cdot \mid x) = \mathrm{N}\left(x, \frac{\sigma^2(e^{\alpha t} - 1)}{2\alpha}\right). \tag{6.12}$$

After an application of Bayes theorem and following Section 3.5 we can define

$$F_{X_t|Y_t}(\cdot \mid y) = \mathrm{N}\left(ye^{-\alpha t} + \mu(1 - e^{-\alpha t}), \frac{\sigma^2(1 - e^{-\alpha t})}{2\alpha}\right). \tag{6.13}$$

With these conditionals we compute (6.1) to get the following transition probability

$$\mathrm{P}_t(x, \cdot) = \mathrm{N}\left(xe^{-\alpha t} + \mu(1 - e^{-\alpha t}), \frac{\sigma^2(1 - e^{-2\alpha t})}{2\alpha}\right).$$

We see that $\mathrm{P}_t(x, \cdot) \to Q_X(\cdot)$ when $t \to \infty$, verifying that the stationary distribution

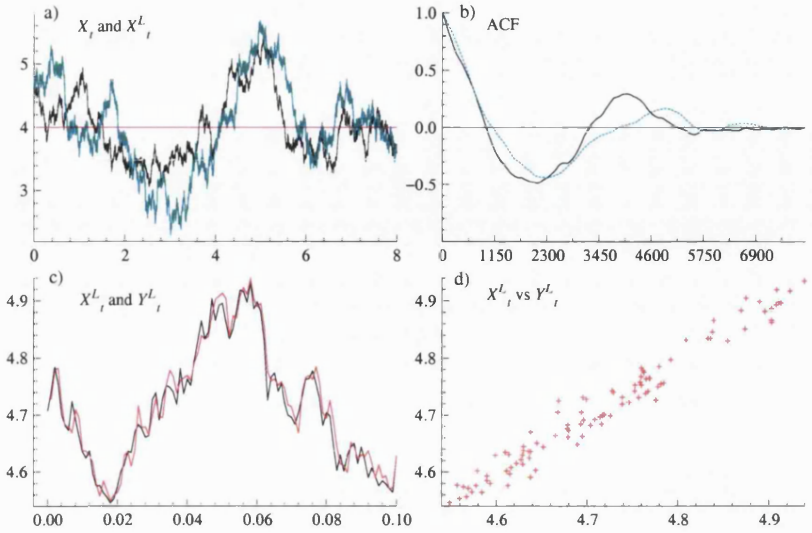


Figure 6.1: Simulation of a mean reverting Gaussian OU process, with parameters $\mu = 4, \sigma^2 = 1, \alpha = 1$. The super-index L indicates that the OU process was simulated using the latent structure (6.12) and (6.13). *a*) Simulated paths: the bold path corresponds to X_t. *b*) The ACF for both simulations: the dotted line corresponds to X^L_t. *c*) The paths on $(0, 0.1]$ corresponding to both the latent and the target process, the latter in bold. Finally *d*) shows the scatter plot of X^L_t vs Y^L_t corresponding to the paths on $(0, 0.1]$.

is also the limiting one. An analog of Proposition 6.1 follows immediately. The drift and diffusion coefficients in this case are given by $\mu(x) = -\alpha(x - \mu)$ and $\sigma(x) = \sigma$ respectively. The Dynkin condition can be also verified exactly as done in equation (6.7). This ensures the representation (up to a suitable specification of the boundary behavior) of a weak solution to the following SDE

$$dX_t = -\alpha(X_t - \mu)\, dt + \sigma\, dW_t,$$

known as the mean reverting Gaussian Ornstein-Uhlenbeck SDE. Figure 6.1 presents some simulations of this process.

Notice that, in this section, the time-dependence function used in the conditionals was chosen by first computing the n-step transition probability in a discrete-time model and then extending this function to a continuous-time setting. This gave us a clear choice of the required time-dependence functions $\phi(\cdot)$ to construct a valid transition probability that satisfied the Chapman-Kolmogorov equations and the Dynkin condition. However, as we mentioned in the introduction of this chapter, this mechanism is not likely to work

for more complicated models where the n-step transition probability is not analytically available. As we will see in the following section, a different technique could be used when the n-step transition is not analytically available.

It is worth mentioning that the Chapman-Kolmogorov equations are not needed once the Dynkin condition has been satisfied. However, obtaining the right time-dependence relations for the conditionals, such that the Dynkin condition is satisfied, is not straightforward task.

From an estimation perspective, it can be argued that our approach to OU processes does not change significantly the dimension of the problem since, both conditional distributions and transition probabilities are normally distributed. Therefore, estimation procedures that use the latent decompositions, such as the EM algorithm, will have the same degree of complexity of those procedures that do not use it, such as the MLE method. However, as we will see in Section 6.6 this is not always the case. It is often easier to work with the conditionals involved in our construction than with the transition probability. Furthermore, there could be cases where transition probabilities are not in a closed form whilst the conditional probabilities are.

6.3 POISSON-GAMMA MODELS

Following Chapter 3, let us assume that $\mathrm{E} = \mathbb{R}_+$, $\mathrm{S} = \mathbb{N}_0$, $\mathcal{T} = \mathbb{R}_+$, λ_x is the Lebesgue measure and λ_y is the counting measure. Suppose we want to construct a stationary stochastic process $\mathrm{X} = \{X_t; t \in \mathbb{R}_+\}$ with invariant measure which has density given by $q_x(x) \propto x^{a-1}e^{-bx}$, $b \geq 0$. Bear in mind that, the invariant density does not necessarily lead to a finite measure, since $q_x(\cdot)$ is not integrable for $b = 0$. For $b > 0$, we use the normalized density, such that $q_x(x) = \mathrm{Ga}(x; a, b)$ is the invariant distribution. In other words, the model can have either proper or improper invariant measures. Clearly, in the case of an infinite invariant measure we will not end up with a stationary process. Applying the second method described in Section 6.1, we fix one of the conditional distributions required in the Gibbs type construction to be

$$f_{Y_t|X_0}(y \mid x) = \mathrm{Po}(y; x\phi_t), \quad \phi_t \geq 0 \quad \forall t > 0, \tag{6.14}$$

where ϕ_t is a time-dependence function to be found. For the moment, we assume that $\phi_t \in C_0$ with

$$C_0 = \left\{ g : [0,\infty) \to \mathrm{R}_+ \mid \lim_{x\to\infty} g(x) = 0 \right\}.$$

That is, ϕ_t is a positive function vanishing at infinity.

An application of Bayes theorem gives the required expression for the complementary conditional required in the Gibbs type construction, for the model at issue we have

$$f_{X_t|Y_t}(x \mid y) = \mathrm{Ga}(x; y+a, \phi_t + b). \tag{6.15}$$

In this case, the transition density for the target process X is given by

$$\begin{aligned} p_t(x',x) &= \sum_{y=0}^{\infty} \mathrm{Ga}(x; y+a, \phi_t+b)\,\mathrm{Po}(y; x'\phi_t) \\ &= \frac{e^{-[\phi_t(x+x')+bx]}}{(\phi_t+b)^{-(a+1)/2}\,\phi_t^{(a-1)/2}} \sqrt{\frac{x}{x'}}^{\,a-1} \mathrm{I}_{a-1}\left(2\sqrt{xx'\phi_t(\phi_t+b)}\right), \end{aligned} \tag{6.16}$$

where $t, x, x' > 0$ and $I_\nu(\cdot)$ denotes the modified Bessel function of the first kind with index ν. See Abramowitz and Stegun [2]. In order to see for which values of ϕ expression (6.16) satisfies the Chapman-Kolmogorov equations (2.5), it is easier to deal with the Laplace transform than with the transition density[2]. Denote the Laplace transform of the random variable Z as $\mathcal{L}_Z(\varsigma) := \mathbb{E}\left[e^{\varsigma Z}\right]$. Remember that if $Z \sim \mathrm{Ga}(a,b)$, then $\mathcal{L}_Z(\varsigma) = (1-b^{-1}\varsigma)^{-a}$ and if $Z \sim \mathrm{Po}(\eta)$, then $\mathcal{L}_Z(\varsigma) = \exp\{\eta(e^\varsigma - 1)\}$. The Laplace transform for the transition (6.16) can be easily found by using the latent decomposition as follows

$$\begin{aligned} \mathcal{L}_{X_t|X_0=x}(\lambda) &= \mathbb{E}\left[\mathcal{L}_{X_t|Y_t}(\lambda) \mid X_0 = x\right] \\ &= \left\{1-(\phi_t+b)^{-1}\lambda\right\}^{-a} \mathcal{L}_{Y_t|X_0}\left(-\ln\left(1-(\phi_t+b)^{-1}\lambda\right)\right) \\ &= \left\{1-(\phi_t+b)^{-1}\lambda\right\}^{-a} \exp\left\{\frac{x\phi_t\lambda}{\phi_t+b-\lambda}\right\}. \end{aligned} \tag{6.17}$$

Notice that

$$\lim_{x\downarrow 0} \mathcal{L}_{X_t|X_0=x}(\lambda) = \left\{1-(\phi_t+b)^{-1}\lambda\right\}^{-a}.$$

Therefore, by inverting the Laplace transform we find that the transition starting at

[2]For construction of processes with densities, such as (6.16), see Feller [34].

zero is given by

$$p_t(x', x) = \text{Ga}(a, \phi_t + b), \quad x' = 0. \tag{6.18}$$

Proposition 6.2. The process $\mathrm{X} = \{X_t; t > 0\}$ with transition densities given by (6.16) and time-dependent function

$$\phi_t = \begin{cases} \frac{b\rho^t}{1-\rho^t} & \text{for } b > 0, \quad 0 < \rho < 1 \\ \frac{1}{ct} & \text{for } b = 0, \quad c > 0, \end{cases}$$

satisfies the Chapman-Kolmogorov equations.

Proof. In terms of Laplace transforms, the Chapman-Kolmogorov equations can be written as follows

$$\mathbb{E}\left[\mathcal{L}_{X_{t+s}|X_s}(\lambda) \mid X_0\right] = \mathcal{L}_{X_{t+s}|X_0}(\lambda). \tag{6.19}$$

Then, in this case

$$\begin{aligned} & \mathbb{E}\left[\mathcal{L}_{X_{t+s}|X_s}(\varsigma) \mid X_0 = x\right] \\ = \; & \left\{1 - (\phi_t + b)^{-1}\lambda\right\}^{-a} \mathcal{L}_{X_s|X_0}\left(\frac{\phi_t \lambda}{\phi_t + b - \lambda}\right) \\ = \; & \left\{1 - \frac{\lambda(\phi_s + \phi_t + b)}{(\phi_t + b)(\phi_s + b)}\right\}^{-a} \exp\left\{\frac{x\lambda\phi_s\phi_t}{(\phi_t + b)(\phi_s + b) - \lambda(\phi_s + \phi_t + b)}\right\}. \end{aligned} \tag{6.20}$$

We need to show that the latter equation is equal to $\mathcal{L}_{X_{t+s}|X_0=x}(\lambda)$.

Case 1: $b = 0$ (improper invariant measure)

If $b = 0$, then from (6.20)

$$\frac{\phi_{t+s}}{\phi_{t+s} - \lambda} = \frac{\phi_s\phi_t}{\phi_s\phi_t - \lambda(\phi_s + \phi_t)},$$

which is equivalent to

$$\phi_{t+s} = \frac{\phi_s\phi_t}{\phi_s + \phi_t}.$$

Assuming that $\phi \neq 0$ the above expression may be written as

$$\frac{1}{\phi_{t+s}} = \frac{1}{\phi_t} + \frac{1}{\phi_s}.$$

The only Lebesgue measurable solution to this equation is given by $\phi_t = 1/ct$, $c > 0$ otherwise the solution is unbounded. See Feller (1966).

Case 2: $b > 0$ **(proper invariant distribution Ga(a, b))**

The intuition used here to get a valid time-dependent function is given by means of the following observation. Using the transition (6.17) we have

$$\mathcal{L}'_{X_t|X_0=x}(0) = \frac{a}{b+\phi_t} + \frac{x\phi_t}{b+\phi_t}.$$

Notice that a process with

$$\mathbb{E}[X_t \mid X_0 = x] = \mu + x\rho^t, \quad 0 < \rho < 1$$

defines a stationary process. Comparing the above conditional moments we get

$$\frac{\phi_t}{b+\phi_t} = \rho^t \Leftrightarrow \phi_t = \frac{b\rho^t}{1-\rho^t}. \tag{6.21}$$

Hence substituting (6.21) in (6.17), equality (6.19) follows immediately. □

Note that Proposition 6.2 leads to two different models, one with a proper invariant measure ($b > 0$) and the other with an improper invariant measure ($b = 0$).

6.4 DIFFUSIONS RELATED TO POISSON-GAMMA MODELS

Using the transition densities (6.16) we can relate (weakly) the target process with some diffusion models. As reviewed in Section 6.2, if we assume that our target process X is standard and that the Dynkin condition is satisfied then the process X can be seen as a version of a diffusion process with homogeneous drift and diffusion coefficients given by (2.18) and (2.19) respectively. See also Karlin & Taylor [53]. Once more we consider the following cases.

Case 1: $b = 0, \phi_t = \frac{1}{ct}$

In this case the conditional moments can be computed via the Laplace transform as

follows:

$$\begin{aligned}
\mathcal{L}'_{X_t|X_0=x}(0) &= act + x \\
\mathcal{L}''_{X_t|X_0=x}(0) &= (act+x)^2 + ct(act+2x) \\
\mathcal{L}^{(4)}_{X_t|X_0=x}(0) &= 24xc^3t^3 + 36x^2c^2t^2 + 12x^3ct + x^4 + 4actx^3 + 44ac^3t^3x \\
&+ 30ac^2t^2x^2 + 24a^2c^3t^3x + a^4c^4t^4 + 6a^3c^4t^4 + 4a^3c^3t^3x + 11a^2c^4t^4 \\
&+ 6a^2c^2t^2x^2 + 6ac^4t^4.
\end{aligned}$$

With these moments it is straightforward to compute the corresponding infinitesimal drift and diffusion coefficients as well as to verify the Dynkin condition.

$$\begin{aligned}
\mu(x) &= \lim_{h\downarrow 0} \frac{1}{h}\left\{\mathcal{L}'_{X_h|X_0=x}(0) - x\right\} = \lim_{h\downarrow 0}\frac{1}{h}\{ach\} = ac \\
\sigma^2(x) &= \lim_{h\downarrow 0}\frac{1}{h}\left\{\mathcal{L}''_{X_h|X_0=x}(0) + x^2 - 2x\mathcal{L}'_{X_h|X_0=x}(0)\right\} \\
&= \lim_{h\downarrow 0}\frac{1}{h}\left\{a^2c^2h^2 + ac^2h^2 + 2cxh\right\} = 2cx.
\end{aligned}$$

For the Dynkin condition, we get

$$\begin{aligned}
&\lim_{h\downarrow 0}\frac{1}{h}\left\{\mathcal{L}^{(4)}_{X_h|X_0=x}(0) - 4x\mathcal{L}^{(3)}_{X_h|X_0=x}(0) + 6x^2\mathcal{L}''_{X_h|X_0=x}(0) - 4x^3\mathcal{L}'_{X_h|X_0=x}(0) + x^4\right\} \\
= &\lim_{h\downarrow 0}\frac{1}{h}\left\{h^2c^2\left(12x^2 + hcx(36a + 12a^2 + 24) + h^2c^2(a^4 + 6a^3 + 11a^2 + 6a)\right)\right\} \\
= &\ 0.
\end{aligned}$$

Therefore, the model constructed using the conditional distributions (6.14) and (6.15) (with $b = 0$ and $\phi_t = 1/ct$) can be seen as the diffusion process that solves the SDE

$$dX_t = ac\,dt + \sqrt{2cX_t}\,dW_t, \quad X_0 = x \geq 0, \tag{6.22}$$

where W_t denotes a one-dimensional Brownian motion.

In this case the transition density (6.27) reduces to

$$p_t(x', x) = \begin{cases} \frac{e^{-\frac{x+x'}{ct}}}{ct}\left(\sqrt{\frac{x}{x'}}\right)^{a-1} \mathrm{I}_{a-1}\left(\frac{2\sqrt{xx'}}{ct}\right) & t > 0, \quad x, x' > 0, \\ \mathrm{Ga}(x; a, 1/ct) & x' = 0. \end{cases} \tag{6.23}$$

Case 2: $b > 0$ and $\phi_t = \frac{b\rho^t}{1-\rho^t}$. Here we have

$$\begin{aligned}
\mathcal{L}'_{X_t|X_0=x}(0) &= \frac{a}{b} + \rho^t(x - \frac{a}{b}) \\
\mathcal{L}''_{X_t|X_0=x}(0) &= \frac{(a + x\rho^t b)^2 - 2\rho^t[a^2 + a + xb(a\rho^t + \rho^t - a)] + \rho^{2t}(a^2 + a) + a}{b^2} \\
\mathcal{L}^{(4)}_{X_t|X_0=x}(0) &= \frac{\rho^{4t}}{b^4}\{x^4b^4 - C + B - A[4xb - a]\} + \frac{\rho^{3t}}{b^4}\{C - 2B + A[12xb - 4a]\} \\
&+ \frac{\rho^{2t}}{b^4}\{B - A[12xb - 6a]\} + \frac{\rho^t}{b^4}\{A[4xb - 4a]\} + aA,
\end{aligned}$$

where $A := a^3 + ba^2 + 11a + 6$, $B := 6x^2b^2[a^2 + 5a + 6]$ and $C := 4x^3b^3[a + 12]$. Similarly, using these moments, the Dynkin condition is satisfied for $p = 4$ and the infinitesimal drift and diffusion coefficients are given by

$$\mu(x) = \ln\left(\frac{1}{\rho}\right)\frac{(a - xb)}{b}, \qquad \sigma(x) = \sqrt{2\frac{x}{b}\ln\left(\frac{1}{\rho}\right)}. \tag{6.24}$$

Note that in contrast to (6.22), a diffusion model with the above coefficients has a proper invariant measure being $\text{Ga}(a, b)$ distributed. This model can be seen, after a suitable re-parametrization as the well-known Cox-Ingersoll-Ross model for interest rates, see Cox et al [24]. This relation will be justified below. In Figure 6.2 some simulations of the diffusion with coefficients (6.24) are presented. Notice that the process X concentrates around a/b since this is the mean for the invariant distribution $\text{Ga}(a, b)$.

As we have mentioned in Chapter 3, one of the issues to consider, while fixing the choice of the conditional distribution $F_{Y_t|X_0}$, is the identifiability of the resulting model as a well-known model. For example, if we associate a Gibbs type construction to a particular diffusion process, then we can use the latent structure to estimate the underlying parameters in the model, for instance by using the methods presented in Section 3.7. So far the choice for both the invariant measure Q_X and the conditional distribution $F_{Y_t|X_0}$, leads to stationary continuous-time models living on the positive real line. The latter features characterize diffusion models in many applications, in particular those related with interest rate models. In what follows we review some well-known diffusion models that can be seen as particular cases of the Poisson-gamma model.

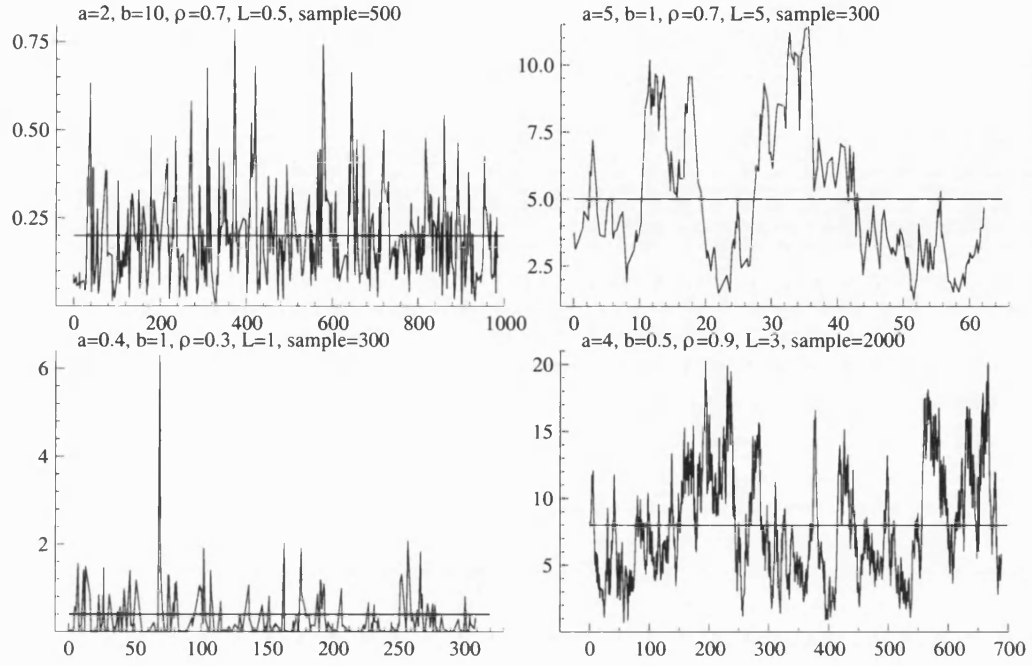


Figure 6.2: Simulations of the positive recurrent process X for the case $b > 0$ in Proposition 6.2. The inter-arrival times are exponentially distributed with rate L. The horizontal line represents a/b, the mean of a gamma random variable. The simulations were done using the Gibbs sampler type representation.

Case 1: Bessel process

The δ-Bessel process $\mathrm{Z} = \{Z_t; t \in \mathbb{R}_+\}$ is defined as the Euclidean distance from the origin of a δ-dimensional Brownian Motion $\delta \geq 0$. That is,

$$Z_t = \sqrt{W_1(t)^2 + W_2(t)^2 + \cdots + W_\delta(t)^2}, \tag{6.25}$$

where $W_i(t)$, $t \geq 0$ are independent one-dimensional Brownian motions. It can be shown that this process can be written as the solution to the following stochastic differential equation

$$dZ_t = \frac{\delta - 1}{2Z_t}\, dt + dB_t, \quad Z_0 = z_0 > 0,$$

where B is the one dimensional Brownian motion independent of Z. See Karlin and Taylor (1981).

Now, if we consider the transformation $X = Z^2$, then the resulting process is given by

$$\begin{aligned} dX_t &= 2Z_t\, dZ_t + dt \\ dX_t &= 2Z_t \left(\frac{\delta - 1}{2Z_t}\, dt + dB_t \right) + dt \\ dX_t &= 2\sqrt{X_t}\, dB_t + \delta\, dt, \quad X_0 = x \geq 0 \end{aligned} \tag{6.26}$$

known as the δ-dimensional squared Bessel processes and denoted by $BESQ^{\delta}(x)$. For this process, the drift and diffusion coefficients satisfy the regularity conditions (2.20) and there exists a linear growth bound. Under these conditions it is possible ensure the existence and uniqueness of a strong solution. See Rogers & Williams [85]. The importance of this process is mainly due to its close relation to other processes with applications in finance such as the CIR model and the geometric Brownian motion. For a good summary and generalizations of $BESQ^{\delta}(x)$ processes we refer to Göing-Jaeschke and Yor [41].

The $BESQ^{\delta}(x)$ process (6.26) can be seen (in a weak sense) as a particular case of the Poisson-gamma[3] model (with $b = 0$), which we have shown has the representation (6.22). The simple re-parametrization $c = 2$ and $a = \delta/2$ establishes the relation. The transition densities for $BESQ^{\delta}$ are given by

$$p_t(x', x) = \begin{cases} \frac{e^{-\frac{x+x'}{2t}}}{2t} \left(\sqrt{\frac{x}{x'}}\right)^{\frac{\delta}{2}-1} \mathrm{I}_{\frac{\delta}{2}-1}\left(\frac{\sqrt{xx'}}{t}\right) & t > 0, \quad x, x' > 0, \\ \mathrm{Ga}(x; \delta/2, 1/2t) & x' = 0, \end{cases} \tag{6.27}$$

see, for instance, Karlin and Taylor [53]. The transition density (6.27) matches the corresponding transition density obtained from (6.16) and (6.18), once the corresponding parameters $b = 0, a = \delta/2, \phi_t = 1/2t$ $(c = 2)$ are substituted.

Notice that with the representation in terms of conditionals, some distributional properties of $BESQ^{\delta}$ can be easily stated in terms of the gamma and Poisson distributions. For instance, the additivity property of squared Bessel processes, that is, the convolution of the laws corresponding to $BESQ^{\delta}(x)$ and $BESQ^{\delta'}(x')$, is the law corresponding to $BESQ^{\delta+\delta'}(x + x')$. See Shiga and Watanabe [94]. This is easily stated in terms of

[3]For the general representation of the $BESQ^{\delta}(x)$ process through a Poisson-Gamma model, we must allow the parameter $a \geq 0$, so the case $\delta = 0$ is also covered.

gamma and Poisson distributions, since

$$\mathrm{Ga}\left(\frac{\delta}{2},\frac{1}{2t}\right) * \mathrm{Ga}\left(\frac{\delta'}{2},\frac{1}{2t}\right) = \mathrm{Ga}\left(\frac{\delta}{2}+\frac{\delta'}{2},\frac{1}{2t}\right)$$

and

$$\mathrm{Po}\left(\frac{x}{2t}\right) * \mathrm{Po}\left(\frac{x'}{2t}\right) = \mathrm{Po}\left(\frac{x+x'}{2t}\right),$$

where $Q * P$ stands for convolution.

The $BESQ^{\delta}(x)$ model serves as an example of a continuous time model that can be represented using the Gibbs type construction but without being stationary. Notice that for this case, $Q_X(dx) = x^{a-1}dx$ is an invariant measure that is not finite. Therefore we can say, by construction, that the $BESQ^{\delta}(x)$ model is Q_X-symmetric though not reversible.

For the squared Bessel process, the value for the parameter δ classifies the boundary points and some stability properties. More precisely, if we notice that the scale measure corresponding to the model (6.26) is given through

$$S'(\eta) = \exp\left\{-\int^{\eta} \frac{\delta}{2z}\, dz\right\},$$

that is,

$$S(\eta) = \begin{cases} \log\eta, & \text{if} \quad \delta = 2 \\ -\eta^{1-\frac{\delta}{2}}, & \text{if} \quad \delta > 2 \\ \eta^{1-\frac{\delta}{2}}, & \text{if} \quad 0 \leq \delta < 2. \end{cases} \tag{6.28}$$

The corresponding speed measure has Lebesgue density given by

$$m(\eta) = \begin{cases} \frac{1}{2}, & \text{if} \quad \delta = 2 \\ \frac{\eta^{\frac{\delta}{2}-1}}{\delta-2}, & \text{if} \quad \delta > 2 \\ -\frac{\eta^{\frac{\delta}{2}-1}}{\delta-2}, & \text{if} \quad 0 \leq \delta < 2. \end{cases} \tag{6.29}$$

Then, using the above quantities it is possible to show that, in particular, 0 is an absorbing point for the $BESQ^0$. This can be done by noticing that $M((0,x]) = \infty$ and $\int_0^x S(\eta)\, m(\eta)\, d\eta < \infty$. Furthermore, for $\delta \geq 3$ the $BESQ^{\delta}$ is transient and for $\delta \leq 2$, it is recurrent. For these and some other properties of the $BESQ^{\delta}$ we

refer to Revuz and Yor [81] and Rogers and Williams [85]. Figure 6.3 shows some simulations of $BESQ^{\delta}(x)$ processes, notice how the transience and recurrence is clear for the corresponding parameter values.

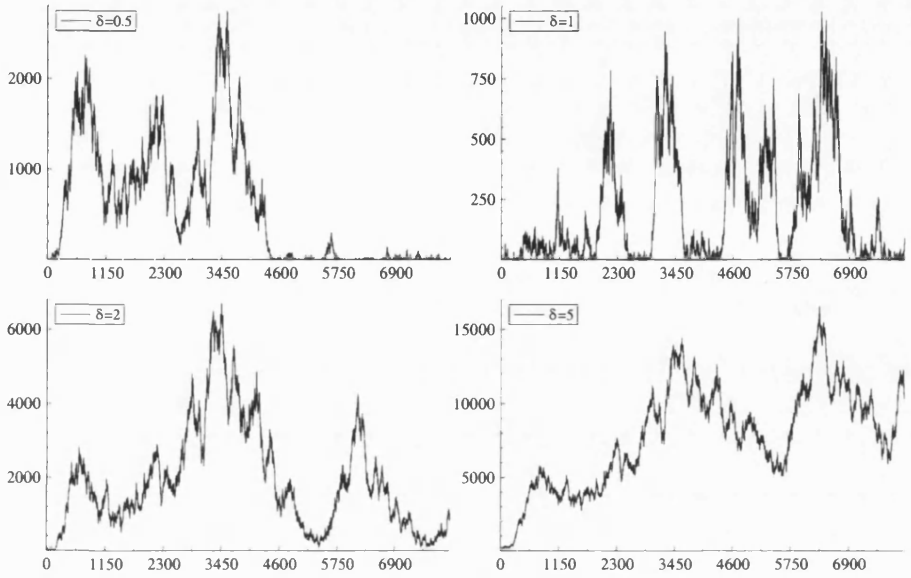


Figure 6.3: Simulations of the squared Bessel process $BESQ^{\delta}(x)$ for different dimensions. For the simulations we have use the Poisson-Gamma representation.

Case 2: Cox-Ingersoll-Ross model

The Cox-Ingersoll-Ross (CIR) family of diffusions, is given by the solutions of

$$dY_t = (\alpha + \beta Y_t)\, dt + \gamma \sqrt{Y_t}\, dW_t, \qquad Y_0 = y \geq 0 \tag{6.30}$$

with[4] $\alpha \geq 0$, $\gamma > 0$, $\beta < 0$. This model has been proposed to model short-term interest rates, see Cox et al [24]. When $\alpha = 0$, SDE (6.30) is also known as Feller's branching diffusion. This family of diffusions is embraced by the Poisson-gamma model, for the case where $b > 0$. A simple re-parametrization establishes the relation

$$\alpha = \frac{\ln\left(\frac{1}{\rho}\right) a}{b} > 0, \quad \beta = \ln(\rho) < 0 \text{ and } \gamma = \sqrt{\frac{2}{b} \ln\left(\frac{1}{\rho}\right)} > 0.$$

[4] $\alpha \in \mathrm{R}$ leads to a more general family not considered here, see Göing-Jaeschke and Yor [41].

The relation between models $BESQ^{\delta}(x)$ and CIR is well-known within the stochastic finance literature. The transition density corresponding to (6.30) can be found by the following scale-time change of the equation (6.26)

$$Y_t = e^{\beta t} X\left(\frac{\gamma^2}{4\beta}(1-e^{-\beta t})\right),$$

with $\delta = 4\alpha/(\gamma^2)$.

In the same way, the transition density corresponding to the diffusion

$$dY_t = \ln\left(\frac{1}{\rho}\right)\frac{(a-Y_t b)}{b}\,dt + \sqrt{2\frac{Y_t}{b}\ln\left(\frac{1}{\rho}\right)}\,dW_t,$$

is given by simply making a time change of the diffusion

$$dX_t = \xi c\,dt + \sqrt{2cX_t}\,dW_t,$$

that is, by applying Itô lemma (2.15) to

$$Y_t = \rho^t X\left(\frac{\rho^{-t}-1}{2b}\right) \tag{6.31}$$

with $c = 2$ and $\xi = a$. Notice that, if we have a random variable $X \sim f_X(x)$ then $Y = \rho^t X \sim f_X(y\rho^{-t})\rho^{-t}$. Thus the transition density of (6.31) is given by

$$p_t^Y(y',y) = \frac{p_{\xi_t}^X(y',\rho^{-t}y)}{\rho^t}, \qquad \xi_t = \frac{\rho^{-t}-1}{2b}$$

where $p_t^X(\cdot,\cdot)$ is given by (6.23). That is

$$p_t^Y(y',y) = \frac{b\exp\left\{-\frac{b(\rho^{-t}y+y')}{\rho^{-t}-1}\right\}}{(\rho^{-t}-1)\rho^{t\frac{a+1}{2}}}\left(\sqrt{\frac{y}{y'}}\right)^{a-1} I_{a-1}\left(\frac{2b\rho^{-t/2}\sqrt{yy'}}{\rho^{-t}-1}\right).$$

This expression agrees with the transition (6.16) obtained using the Poisson-gamma based model.

Within the financial literature, modelling of interest rates commonly requires processes with stationary behavior other that the one presented by the gamma distributions. This is mainly because underlying stationary distributions might have heavier tails when describing short interest rates. This in turn leads to the requirement of other

stationary continuous-time models.

A simple change in the Poisson-gamma model described in this section will result in a diffusion with invariant distribution being inverse gamma. In order to see this, let us assume that the invariant measure for the target process X is $Q_X = \text{Iga}(a, b)$, $a > 0, b > 0$, that is $X = 1/Y$, where $Y \sim \text{Ga}(a, b)$. As for the Poisson-gamma models, we assume

$$f_{Y_t|X_0}(y \mid x) = \text{Po}(y; \phi_t/x). \tag{6.32}$$

Here, we use the same time-dependent function (for the proper case, $b > 0$) given by $\phi_t = b\rho^t/(1-\rho^t)$. The other conditional needed for the Gibbs type construction is fixed by

$$f_{X_t|Y_t}(x \mid y) = \text{Iga}(x; y + a, \phi_t + b), \tag{6.33}$$

If we denote the transition density (6.16) by $p_t^{Ga}(x', x)$, then the transition corresponding to the conditionals (6.32) and (6.33) is given by

$$\begin{aligned} p_t^{Iga}(x', x) &= \sum_{y=0}^{\infty} \text{Iga}(x; y + a, \phi_t + b)\, \text{Po}(y; \phi_t/x') \\ &= \frac{p_t^{Ga}(1/x', 1/x)}{x^2}. \end{aligned}$$

Which is up to say that both, the transition density and the invariant measure, have been transformed. In this case, for $a \geq 4$ we verified the Dynkin condition and computed the infinitesimal coefficients. The resulting diffusion can be written as the weak solution to

$$dX_t = \left\{ \frac{\ln(\rho)}{b}(a-2)X_t^2 - \ln(\rho)X_t \right\} dt + \sqrt{\frac{2{X_t}^3}{b} \ln\left(\frac{1}{\rho}\right)}\, dW_t, \;\; X_0 = x > 0. \tag{6.34}$$

If we let $\rho = e^{-1}$, then the above SDE rewrites as

$$dX_t = \left\{ X_t - \frac{(a-2)}{b} X_t^2 \right\} dt + \sqrt{\frac{2}{b} X_t^3}\, dW_t, \;\; X_0 = x > 0. \tag{6.35}$$

Other diffusion processes may be represented with the same time-dependence function ϕ by means of transforming the gamma random variable.

In addition to the relation between the transition densities corresponding to the Poisson-

gamma model and the presented diffusions, one may ask which is the relation or interpretability for the underlying conditionals. An answer to these questions may be motivated by one of the Ray-Knight theorems for local times of Brownian motion.

If we denote $\{L_t^a : t \geq 0, a \in \mathrm{R}\}$ as the local time in a of the Brownian motion W, then one of the versions of the Ray-Knight theorem states that if $\tau := \inf\{t : W_t = -1\}$, then the process $\{L_\tau^x; x \geq -1\}$ is a time-inhomogeneous diffusion. The latter behaves like a squared Bessel process of dimension 2 for $-1 \leq x \leq 0$, and of dimension 0 for $x > 0$. A clear proof of this result using the excursion point process is given by Rogers [83]. With the arguments therein, it is possible to see that the local time L_τ^b, with $b = a + \xi \in (a, 0)$ and $a \in (-1, 0)$, (for instance when $-1 \leq x \leq 0$) can be seen as the sum of a $\mathrm{Po}(x_a/2\xi) + 1$ number of independent $\mathrm{Ga}(1, 1/2\xi)$ random variables. Where x_a is the value taken by L_t^a. Therefore, it is easy to see that the transition density underlying to the local time process $(L_\tau^x)_{x \geq -1}$, for $-1 \leq x \leq 0$, is the same as the one that corresponds to a squared Bessel process of dimension 2.[5]

6.5 POISSON-BASED MODELS

For the methods presented so far we had the complete specification of the model. That is, we had either both conditionals or one conditional and the knowledge of the required invariant measure. In what follows, this complete specification is relaxed, in the sense that we only fix the conditional distributions $F_{Y|X}$ and $F_{X|Y}$. By virtue of the alternative construction presented in Section 3.6, a stationary distribution exists, provided the underlying conditionals are compatible.

Once more, we fix the following conditional

$$f_{Y|X}(y \mid x) = \mathrm{Po}(y; x\phi_t), \quad \phi_t \geq 0 \quad \forall t > 0$$

and assume that the other conditional is given through the following mechanism

$$X_t = V_{0t} + \sum_{i=1}^{Y_t} V_{it} \tag{6.36}$$

Where $Y_t \mid X_0 \sim \mathrm{Po}(x\phi_t)$, $V_{0t} \sim G_t$ and $V_{1t}, V_{2t}, \ldots$ is a sequence of IID random

[5] We thank Prof. Chris Rogers for providing us with this interpretation.

variables on R_+ having distribution F_t, $t \geq 0$, independent of Y_t and V_{0t}.

In this case, the Laplace transform of the transition probability corresponding to a target process X is given as follows

$$\begin{aligned}
\mathcal{L}_{X_t|X_0=x'}(\lambda) &= \mathcal{L}_{V_{0t}}(\lambda)\, \mathbb{E}\left[\mathbb{E}\left[e^{\lambda \sum_{i=1}^{y} V_{it}}\right] \mid X_0 = x'\right] \\
&= \mathcal{L}_{V_{0t}}(\lambda)\, \mathbb{E}\left[\mathcal{L}_{V_{1t}}(\lambda)^{Y_t} \mid X_0 = x'\right] \qquad (6.37) \\
&= \mathcal{L}_{V_{0t}}(\lambda) \exp\left\{x' \phi_t \left(\mathcal{L}_{V_{1t}}(\lambda) - 1\right)\right\}. \qquad (6.38)
\end{aligned}$$

Equation (6.37) follows from the independence of $V_{1t}, V_{2t}, \ldots$ and Y_t. Notice that, with the above assumptions, the distribution $F_{X|Y}(x \mid y)$ is given via its Laplace transform, namely

$$\begin{aligned}
\mathcal{L}_{X_t|Y_t=y}(\lambda) &= \mathbb{E}\left[\exp\left\{\lambda\left(V_{0t} + \sum_{i=1}^{y} V_{it}\right)\right\}\right] \\
&= \mathcal{L}_{V_{0t}}(\lambda)\left[\mathcal{L}_{V_{1t}}(\lambda)\right]^y.
\end{aligned}$$

In the approach taken here, the complete specification for the invariant measure Q_X is not immediate, since we have started by fixing the conditional distributions and assumed the independence structure (3.17).

Proposition 6.3. The process $\{X(t); t \in \mathcal{T}\}$ constructed using the transition probability with Laplace transform (6.38) satisfies the Chapman-Kolmogorov equations if and only if the function

$$R_t(\lambda) := \phi_t \left(\mathcal{L}_{V_{1t}}(\lambda) - 1\right) \qquad (6.39)$$

satisfy $R_s \circ R_t = R_{t+s}$ and

$$\mathcal{L}_{V_0(t+s)}(\lambda) = \mathcal{L}_{V_0(t)}(\lambda)\, \mathcal{L}_{V_0(s)}(R_t(\lambda)). \qquad (6.40)$$

Proof. First note that, $\mathcal{L}_{X_t|X_0}(0) = 1$ since $\mathcal{L}_{V_{0t}}(0) = \mathcal{L}_{V_{1t}}(0) = 1$, therefore the transition semigroup is honest on $\mathbb{R}_+$. Hence, we need

$$\mathcal{L}_{X_{t+s}|X_0=x}(\lambda) = \mathbb{E}\left[\mathcal{L}_{X_{t+s}|X_s}(\lambda) \mid X_0 = x\right]$$

which is satisfied if and only if

$$\begin{aligned}\mathcal{L}_{V_{0(t+s)}}(\lambda)e^{\{xR_{t+s}(\lambda)\}} &= \mathcal{L}_{V_{0t}}(\lambda)\,\mathbb{E}\left[e^{\{X_sR_t(\lambda)\}} \mid X_0 = x\right] \\ &= \mathcal{L}_{V_{0t}}(\lambda)\,\mathcal{L}_{V_{0s}}(R_t(\lambda))\,e^{\{xR_s(R_t(\lambda))\}}.\end{aligned}$$

Then, the required conditions, (6.39) and (6.40), follow. □

Let us assume that $R_s \circ R_t = R_{t+s}$ is satisfied. If we fix R_t, then we are interested in identifying the random variables V_{it} such that the function ϕ does not depend on λ. Consider the following example.

Example 6.1. Let V_0 be a degenerate random variable, and assume

$$R_t(\lambda) = \frac{\lambda e^t}{1+\lambda(e^t-1)},$$

clearly $R_s \circ R_t = R_{t+s}$. Hence, if we take $\phi_t = e^{-t}$, then,

$$\mathcal{L}_{V_t}(\lambda) = \frac{\lambda}{1+\lambda(e^t-1)} + 1.$$

It is easy to prove that the above function is completely monotone in λ and that $\mathcal{L}_{V_t}(0^+) = 1$. Therefore, by means of Bernstein's Theorem $\mathcal{L}_{V_t}(\cdot)$ is a well-defined laplace transform. Using the inversion formula for Laplace transform we can see that

$$f_{V_t}(x) = \frac{\exp\left\{\frac{x}{e^{-t}-1}\right\}}{(e^{-t}-1)^2} + \frac{\delta(x)e^{-t}}{e^{-t}-1},$$

where $\delta(\cdot)$ denotes the Dirac delta measure at zero. ○

In general, for a fixed R_t satisfying (6.39), $f_{V_t}(x) = (\phi_t)^{-1}\mathcal{L}^{-1}(R_t(\lambda)) + \delta(x)$, where $\mathcal{L}^{-1}(\cdot)$, denotes the inverse Laplace transform. Using the Binomial theorem it is possible to expand the Laplace transform for the convolution in (6.36). That is, denoting this convolution with Z, we have

$$\begin{aligned}f_Z(x) &= \mathcal{L}^{-1}\left[\left\{\frac{R_t(\lambda)}{\phi_t}+1\right\}^y\right] \\ &= \sum_{k=0}^{y}\binom{y}{k}\phi_t^{y-k}\,\mathcal{L}^{-1}\left\{R_t(\lambda)^{y-k}\right\}. \qquad (6.41)\end{aligned}$$

Notice that (6.41) might need to be normalized. In example (6.1) it is clear that the V's are not closed under convolution, however it can be seen that the y-fold convolution belongs to the class of generalized gamma convolutions. See Bondesson [19]. In this case the invariant distribution is found to be $Q_X(\cdot) = \text{Exp}(1)$.

Example 6.2. Assume $V_{0t} \sim \text{Ga}(a, \phi_t + b)$ and $V_{1t} \sim \text{Ga}(1, \phi_t + b)$, $a > 0$, $b > 0$ then,

$$\mathcal{L}_{X_t|X_0}(\lambda) = \left\{1 - (\phi_t + b)^{-1}\lambda\right\}^{-a} \exp\left\{\frac{x_0\phi_t\lambda}{\phi_t + b - \lambda}\right\}, \tag{6.42}$$

with

$$R_t(\lambda) = \frac{\phi_t\lambda}{\phi_t + b - \lambda} \quad \text{and} \quad \mathcal{L}_{V_{0t}}(\lambda) = \left\{1 - (\phi_t + b)^{-1}\lambda\right\}^{-a}. \tag{6.43}$$

We have that $R_{t+s}(\lambda) = R_s(R_t(\lambda))$ if and only if

$$\phi_{t+s} = \frac{\phi_s\phi_t}{\phi_s + \phi_t + b}. \tag{6.44}$$

An example of a solution to (6.44) is given in Proposition 6.2 by $\phi_t = \frac{b\rho^t}{1-\rho^t}$ for $0 < \rho < 1$. Hence

$$R_t(\lambda) = \frac{\lambda b\rho^t}{b - \lambda(1 - \rho^t)}.$$

∘

The main counterpart of the construction presented in this section is that from the outset we do not know which is the invariant measure, however this can be found by marginalizing the joint distribution given by (3.37).

6.6 AN ESTIMATION EXAMPLE

In this section we proceed as in Sections 4.8-4.10. As in the case of discrete-time models, in the continuous-time setting is also possible to use the latent decomposition in estimation procedures.

Let us consider the stationary ($b > 0$) Poisson-gamma model, described in Section 6.3. For this model, the transition density simplifies as follows

$$p_t(x_0, x_t) = \frac{b\exp\left\{-b\left[\frac{(\rho^t x_0 + x_t)}{1-\rho^t}\right]\right\}\sqrt{\frac{x_t}{x_0}}^{a-1}\text{I}_{a-1}\left(\frac{2b\rho^{t/2}\sqrt{x_0 x_t}}{1-\rho^t}\right)}{(1-\rho^t)(\rho^t)^{(a-1)/2}}.$$

where $a, b, t > 0$ and $0 < \rho < 1$.

Following the methods described in Section 3.7, given a sample $\mathbf{x} = (x_{t_1}, x_{t_2}, \ldots, x_{t_T})$, the likelihood is given by

$$L_{\mathbf{x}}(\theta) = q_X(x_{t_1}) \prod_{i=1}^{T-1} p_{(t_{i+1}-t_i)}(x_{t_i}, x_{t_{i+1}}).$$

For the Poisson-gamma stationary model, we have $q_X(x_{t_1}) = \mathrm{Ga}(x_{t_1}; a, b)$. In general, the time between observations can be included in the estimation procedure. That is, if the observations are not taken in time intervals equally spaced, then we can associate to our target process the time vector $\tau = (\tau_2, \tau_3, \ldots, \tau_T)$, where $\tau_i := t_i - t_{i-1}$, $i = 2, \ldots, T$.[6] These sampling intervals can be set deterministically or through a random distribution. Strictly speaking, in order to have an asymptotically efficient estimator, the above likelihood would have to be modified to include the density corresponding to the time-information distribution. For more on estimation of continuous-time models based on random sampling, we refer to Yacine and Mykland [109].

Analytical MLE of the log-likelihood l_θ is not available, since

$$\frac{\partial \log(I_a(x))}{\partial x} = \frac{I_{a+1}(x)}{I_a(x)} + \frac{a}{x}.$$

This implies that the the scores, underlying to the Poisson-gamma model, depend on b and ρ, through the argument of Bessel functions and therefore analytic solutions to $\partial l_\theta = 0$ with respect to $\theta = (a, b, \rho)$ are not generally available. However, a numeric approach similar to the one implemented in Section 4.8 (using the BFGS algorithm) can be taken. As we have mentioned in Section 4.8, the BFGS algorithm can use numerical or analytical gradient computations in order to find the direction towards the maximum. For the sake of exposition, let us assume that $a = 1$, that is we consider the case where the invariant distribution is exponentially distributed with mean $1/b$. Hence, the remaining parameters to be estimated are $\theta = (b, \rho)$. Under this assumption

[6] Examples of data not equally spaced are "tick by tick" financial data.

the components of the gradient for the complete likelihood are given by

$$\begin{aligned}\frac{\partial l_\theta}{\partial b} &= \frac{T}{b} - x_{t_1} - \sum_{i=1}^{T-1}\left\{\frac{\rho^{\tau_{i+1}}x_{t_i} + x_{t_{i+1}}}{1-\rho^{\tau_{i+1}}}\right\} \\ &+ \sum_{i=1}^{T-1} 2R_0^{\diamond}\left(\frac{2b\rho^{\tau_{i+1}/2}\sqrt{x_{t_i}x_{t_{i+1}}}}{1-\rho^{\tau_{i+1}}}\right)\frac{\rho^{\tau_{i+1}/2}\sqrt{x_{t_i}x_{t_{i+1}}}}{1-\rho^{\tau_{i+1}}},\end{aligned}$$

where $R_a^{\diamond}(x) = \mathrm{I}_{a+1}(x)/\mathrm{I}_a(x)$, and

$$\begin{aligned}\frac{\partial l_\theta}{\partial \rho} &= \sum_{i=0}^{T-1}\left\{\frac{\tau_{i+1}\rho^{\tau_{i+1}}}{\rho(1-\rho^{\tau_{i+1}})} - \frac{b(x_{t_i}+x_{t_{i+1}})\tau_{i+1}\rho^{\tau_{i+1}}}{\rho(1-\rho^{\tau_{i+1}})^2}\right\} \\ &+ \sum_{i=0}^{T-1} R_0^{\diamond}\left(\frac{2b\rho^{\tau_{i+1}/2}\sqrt{x_{t_i}x_{t_{i+1}}}}{1-\rho^{\tau_{i+1}}}\right)\frac{b\tau_{i+1}(\rho^{\frac{\tau_{i+1}}{2}} + \rho^{\frac{3\tau_{i+1}}{2}})\sqrt{x_{t_i}x_{t_{i+1}}}}{\rho(1-\rho^{\tau_{i+1}})^2}\end{aligned}$$

and $\tau_i = 1$ for $i = 2, \ldots, T$. In this section, we will use the numeric scores in the BFGS algorithm to numerically implement the MLE method. However, we have displayed the above analytical scores for comparison porpoises in Theorem 6.1 stated below.

Instead of using the BFGS maximization routine, the estimation procedure can be done using an EM algorithm. For this example, just as in the case presented in Section 4.9 and Section 4.10, the distribution of $\{\mathbf{Y} \mid \mathbf{X}\}$ is required. For the case of models constructed using the Gibbs sampler type scheme described in Section 3, we saw that such distribution can be decomposed in conditional independent random variables $\{Y_{t_{i+1}} \mid X_{t_{i+1}}, X_{t_i}\}$. For the general case where $a > 0$, this distribution has density proportional to

$$\begin{aligned}f(y_{t_{i+1}} \mid x_{t_{i+1}}, x_{t_i}) &\propto \mathrm{Ga}(x_{t_{i+1}}; y_{t_{i+1}} + a, \phi(\tau_{i+1}) + b)\,\mathrm{Po}(y_{t_{i+1}}; \phi(\tau_{i+1})x_{t_i}) \\ &\propto \frac{\left\{\frac{x_{t_{i+1}}x_{t_i}b\phi(\tau_{i+1})}{1-\rho^{\tau_{i+1}}}\right\}^{y_{t_{i+1}}}}{y_{t_{i+1}}!\Gamma(y_{t_{i+1}}+a)}. \qquad (6.45)\end{aligned}$$

The above quantity can be normalized by means of the following result

$$\sum_{i=0}^{\infty}\frac{z^i}{i!\Gamma(i+a)} = I_{a-1}(2\sqrt{z})z^{\frac{1-a}{2}}.$$

Therefore the required density is given by

$$f(y_{t_{i+1}} \mid x_{t_{i+1}}, x_{t_i}) = \frac{\left\{\frac{x_{t_{i+1}} x_{t_i} b^2 \rho^{\tau_{i+1}}}{(1-\rho^{\tau_{i+1}})^2}\right\}^{y_{t_{i+1}} + \frac{a-1}{2}}}{y_{t_{i+1}}! \Gamma(y_{t_{i+1}} + a)\, I_0\left[\frac{2b\rho^{\frac{\tau_{i+1}}{2}} \sqrt{x_{t_{i+1}} x_{t_i}}}{1-\rho^{\tau_{i+1}}}\right]}. \tag{6.46}$$

For the particular case when $a = 1$, the augmented likelihood is given by

$$L^{aug}_{\mathbf{x},\mathbf{y}}(\theta) = \mathrm{Ga}(x_{t_1}; 1, b) \prod_{i=1}^{T-1} \mathrm{Ga}\left(x_{t_{i+1}}; y_{t_{i+1}} + 1, \phi(\tau_{i+1}) + b\right) \mathrm{Po}(y_{t_{i+1}}; \phi(\tau_{i+1}) x_{t_i}),$$

with corresponding scores given by

$$\frac{\partial l^{aug}_{\mathbf{x},\mathbf{y}}(\theta)}{\partial b} = \frac{1}{b}\left(T + 2\sum_{i=1}^{T-1} y_{t_{i+1}}\right) - x_{t_1} - \sum_{i=1}^{T-1} \frac{\rho^{\tau_{i+1}} x_{t_i} + x_{t_{i+1}}}{1 - \rho^{\tau_{i+1}}} \tag{6.47}$$

and

$$\begin{aligned}\frac{\partial l^{aug}_{\mathbf{x},\mathbf{y}}(\theta)}{\partial \rho} = &\sum_{i=1}^{T-1} \frac{(y_{t_{i+1}} + 1)\tau_{i+1}\rho^{\tau_{i+1}-1}}{1 - \rho^{\tau_{i+1}}} - \frac{b x_{t_{i+1}} \tau_{i+1} \rho^{\tau_{i+1}-1}}{(1 - \rho^{\tau_{i+1}})^2} \\ &+ y_{t_{i+1}}\left(\frac{\tau_{i+1}}{\rho} + \frac{\tau_{i+1}\rho^{\tau_{i+1}-1}}{1 - \rho^{\tau_{i+1}}}\right) - \frac{b x_{t_i} \tau_{i+1} \rho^{\tau_{i+1}-1}}{(1 - \rho^{\tau_{i+1}})^2}.\end{aligned} \tag{6.48}$$

Equating expression (6.47) to zero and solving for b we get the following estimator

$$\hat{b}_{aug} = \frac{T + 2\sum_{i=1}^{T-1} y_{t_{i+1}}}{x_{t_1} + \sum_{i=1}^{T-1} \frac{x_{t_{i+1}} + x_{t_i}}{1-\rho^{\tau_{i+1}}}}.$$

In general, the estimator for ρ based on the augmented likelihood, is not directly available. However, if we assume that the observations $\mathbf{x}$ are uniformly spaced, $\tau_i = 1$ for all $i = 2, \ldots, T$, then an estimate for ρ is given by solving the following quadratic equation

$$\rho^2\left(1 - T - \sum_{i=1}^{T-1} y_{t_{i+1}}\right) - \rho\left[b\left\{\sum_{i=1}^{T-1}(x_{t_{i+1}} + x_{t_i})\right\} - T + 1\right] + \sum_{i=1}^{T-1} y_{t_{i+1}} = 0,$$

which have real solutions if the following conditions is satisfied

$$\left|\left[b\left\{\sum_{i=1}^{T-1}(x_{t_{i+1}}+x_{t_i})\right\}-T+1\right]\right| \geq \sqrt{4\left(1-T-\sum_{i=1}^{T-1}y_{t_{i+1}}\right)\sum_{i=1}^{T-1}y_{t_{i+1}}}. \tag{6.49}$$

At this point we have, at least, two alternatives to MLE in order to estimate the parameters in the Poisson-gamma stationary model: a MCEM scheme described in Section 3.7, in which we need to simulate from the latent vector $\mathbf{Y} \mid \mathbf{X}$ or an EM method where the E-step is obtained analytically. The latter method can be simplified using an argument similar to the one described in Theorem 4.2 as we see in the following theorem, considered only for the particular case when $a = 1$.

Theorem 6.1. Let $\mathbf{x} = (x_{t_1}, x_{t_2}, \ldots, x_{t_T})$ be a sample from the stationary Poisson-gamma model (6.16) with $b > 0$ and $a = 1$, then

$$\mathbb{E}_{\theta_{(j)}}\left[\nabla l^{aug}_{\mathbf{x},\mathbf{y}}(\theta)\right] = \nabla l_{\mathbf{x}}(\theta).$$

Proof. We proceed as in the proof of Theorem 4.2. First, let us notice that for $a = 1$ we have

$$\mathbb{E}\left[Y_{t_{i+1}} \mid x_{t_{i+1}}, x_{t_i}\right] = \frac{b\rho^{\frac{\tau_{i+1}}{2}}\sqrt{x_{t_{i+1}}x_{t_i}}}{1-\rho^{\tau_{i+1}}} R_0^{\diamond}\left(\frac{2b\rho^{\frac{\tau_{i+1}}{2}}\sqrt{x_{t_{i+1}}x_{t_i}}}{1-\rho^{\tau_{i+1}}}\right). \tag{6.50}$$

Therefore, a component-wise substitution of this conditional moment in the gradient components (6.47-6.48) leads to the result. That is

$$\begin{aligned}\mathbb{E}\left[\frac{\partial l^{aug}_{\mathbf{x},\mathbf{y}}(\theta)}{\partial b} \mid \mathbf{x}\right] &= \frac{1}{b}\left\{T+2\sum_{i=1}^{T-1}\mathbb{E}[y_{t_{i+1}} \mid x_{t_{i+1}}, x_{t_i}]\right\} - x_{t_1} - \sum_{i=1}^{T-1}\frac{\rho^{\tau_{i+1}}x_{t_i}+x_{t_{i+1}}}{1-\rho^{\tau_{i+1}}} \\ &= \frac{\partial l(\theta)}{\partial b},\end{aligned}$$

$$\begin{aligned}\mathbb{E}\left[\frac{\partial l^{aug}_{\mathbf{x},\mathbf{y}}(\theta)}{\partial \rho} \mid \mathbf{x}\right] = \sum_{i=1}^{T-1} &\frac{(\mathbb{E}[y_{t_{i+1}} \mid x_{t_{i+1}}, x_{t_i}]+1)\tau_{i+1}\rho^{\tau_{i+1}-1}}{1-\rho^{\tau_{i+1}}} - \frac{bx_{t_{i+1}}\tau_{i+1}\rho^{\tau_{i+1}-1}}{(1-\rho^{\tau_{i+1}})^2} \\ &+ \mathbb{E}[y_{t_{i+1}} \mid x_{t_{i+1}}, x_{t_i}]\left(\frac{\tau_{i+1}}{\rho}+\frac{\tau_{i+1}\rho^{\tau_{i+1}-1}}{1-\rho^{\tau_{i+1}}}\right) - \frac{bx_{t_i}\tau_{i+1}\rho^{\tau_{i+1}-1}}{(1-\rho^{\tau_{i+1}})^2} \\ &= \frac{\partial l(\theta)}{\partial \rho}.\end{aligned}$$

□

In the same way as in Theorem 4.2, the above theorem allows us to estimate the parameter vector $\theta = (b, \rho)$, by iteratively maximizing the likelihood (component-wise). In this case, given an initial value ρ_0, the EM iterations translate to individually solve

$$\begin{aligned}
\text{Iteration 1:} \quad & \frac{\partial l_{\mathbf{x}}(b, \rho_0)}{\partial b}\Big|_{b=b_1} = 0, \qquad \frac{\partial l_{\mathbf{x}}(b_1, \rho)}{\partial \rho}\Big|_{\rho=\rho_1} = 0 \\
\text{Iteration 2:} \quad & \frac{\partial l_{\mathbf{x}}(b, \rho_1)}{\partial b}\Big|_{b=b_2} = 0, \qquad \frac{\partial l_{\mathbf{x}}(b_2, \rho)}{\partial \rho}\Big|_{\rho=\rho_2} = 0 \\
& \qquad\qquad\qquad\qquad \vdots \qquad\quad \vdots \\
\text{Iteration J:} \quad & \frac{\partial l_{\mathbf{x}}(b, \rho_{J-1})}{\partial b}\Big|_{b=b_J} = 0, \qquad \frac{\partial l_{\mathbf{x}}(b_J, \rho)}{\partial \rho}\Big|_{\rho=\rho_J} = 0.
\end{aligned} \tag{6.51}$$

The above iterations are always moving towards the maximum due to property (3.42), a property that is valid for all –non-Monte Carlo– based EM algorithms.

Whether to implement a MCEM scheme or to directly apply an EM algorithm (via Theorem 6.1) is the tradeoff between simulating from the distribution (6.46) or numerically solving the equations (6.51). The choice would depend on the computational capabilities to evaluate the Bessel functions involved. Under MCEM scheme the estimates based on augmented data $(\hat{b}^{aug}, \hat{\rho}^{aug})$ do not depend on Bessel functions. However, in order to simulate from the distribution (6.46) some evaluations of Bessel functions might be required, although this can be avoided by simulating via a Metropolis-Hastings algorithm. On the other hand, numeric maximization (for example via BFGS) of the non-augmented likelihood depends on a large number of evaluations of Bessel functions. The difference between MLE and the sequential MLE induced via EM (Theorem 6.1) is that the latter ensures the convergence to the optimum, whereas the numeric MLE, based on algorithms such as BFGS, highly depends on the initial values. In what follows, we analyze an example based on simulated data.

Let us consider two data sets, $\mathbf{x}$ and $\mathbf{x}^{\tau}$, of size T=1000 after a burn in period of 5000 iterations. The simulations were obtained via the representation through the conditionals (6.14-6.15) with $a = 1, b = 3$ and $\rho = 0.7$. For the data set $\mathbf{x}$, we assumed equally spaced data, that is $\tau_i = 1$ for all $i = 2, \ldots, T$. For the data set $\mathbf{x}^{\tau}$, the data were generated at exponential times with intensity parameter $\lambda = 0.5$. For the sake of comparison, we aim to estimate the parameters with two methods; the numeric MLE[7],

[7]For the data set $\mathbf{x}^{\tau}$, we have used what Yacine and Mykland [109] define as PFML (Pretend Fixed Maximum Likelihood) estimators, that is we act as if the sampling intervals are all equal. However, for this data set we have re-scaled the time by $\tau_i^* = \mathbb{E}[Z]\tau_i$ for $i = 2, \ldots, T$ where $Z \sim \text{Exp}(0.5)$.

based on the BFGS algorithm and the MCEM method where the required simulations, from the distribution with density (6.46), are done using the inverse CDF method. The simulated data together with their corresponding ACF's are displayed in Figure 6.4.

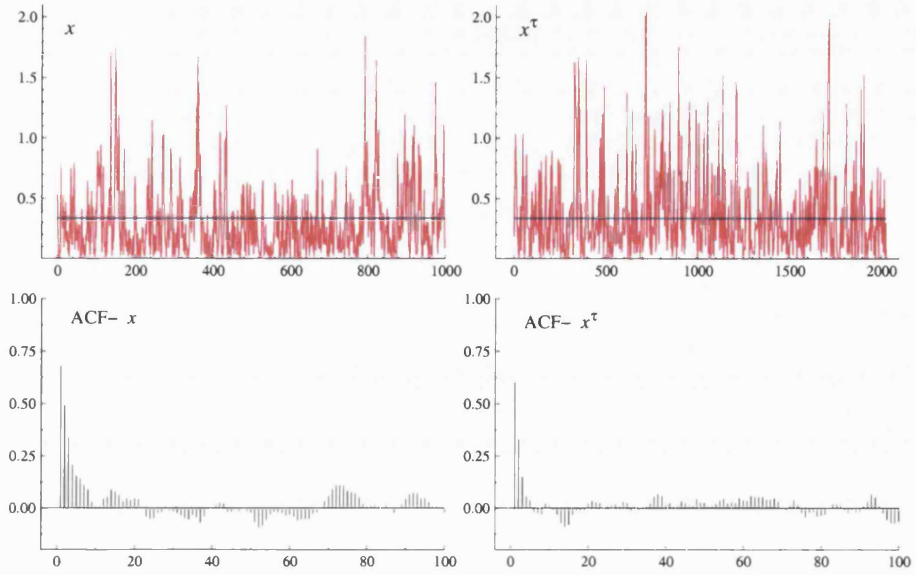


Figure 6.4: Simulated data and respective ACF's from the Poisson-gamma model with parameters $a = 1, b = 3$ and $\rho = 0.7$. The simulations were performed under two schemes: $\mathbf{x}$ denotes an equally-spaced sample and $\mathbf{x}^\tau$ denotes an exponentially-distributed sample (with intensity parameter $\lambda = 0.5$).

Remember that the MCEM differs from the crude EM in that the E-step is approximated by

$$\hat{Q}(\theta \mid \theta_{(j)}, \mathbf{x}) = \frac{1}{m} \sum_{k=1}^{m} \log(L^{aug}_{\mathbf{x},\mathbf{y}^{(k)}}(\theta)), \tag{6.52}$$

where for $k = 1, \ldots, m$, $\mathbf{y}^{(k)}$ denote the random vectors, simulated (component-wise) from (6.46).

Table 6.1 shows the behavior of the above estimate as the sample size m increases. We observed that only a few simulations are required in order to get a relatively good estimation. With simulations larger than 30, the resulting estimates, obtained from the MCEM, did not show a significant improvement. Being the latter our main objective, we initially implemented the MCEM fixing $m = 30$. For the data set $\mathbf{x}$, 100 iterations

Simulations (m)	$\hat{Q}(\theta \mid \theta, \mathbf{x})$	$\hat{Q}(\theta \mid \theta, \mathbf{x}^\tau)$
(Model)	-731.49	-787.71
1	-688.53	-921.40
5	-711.89	-590.12
10	-708.63	-723.58
30	-717.68	-782.63
100	-721.03	-781.31
1000	-731.94	-786.10

Table 6.1: Monte Carlo approximation for the E-step, evaluated at the true parameter-value.

took 1.27 min. whereas for the data set $\mathbf{x}^\tau$ the time needed for 100 iterations was 1.57 min. Improvement in computational-time was attained by allowing the number of simulations from $\mathbf{Y} \mid \mathbf{X}$ to vary within each iteration of the MCEM algorithm. We explore several examples to ensure that the resulting estimates remain satisfactory.

If we fix the initial number of simulations to 30 (so that a good approximation is provided at the outset) and then decrease it by half within each iteration, with a minimum of 5 simulations per iteration, then 100 MCEM iterations corresponding to the data set $\mathbf{x}$ took 20.4 sec. whereas for the data set $\mathbf{x}^\tau$ took 27.4 sec.[8]

Tables 6.3 and 6.4 show the MCEM iterations for the uniformly spaced data set $\mathbf{x}$ and the exponentially spaced data set $\mathbf{x}^\tau$ respectively. In this case we have displayed the results for the case where the number of MC simulations decreases within each iteration. For the randomly spaced data set the parameter ρ, which represents the correlation of the model, is not as close to the theoretical value as in the case of the uniformly spaced data set. This is mainly due to the fact that we have ignored the randomness of the time-gap between observations. Ways to correct this issue are studied in Yacine and Mykland [109].

The results for the numeric maximization of the likelihood via the BFGS algorithm (numeric gradient) are displayed in Table 6.2 only for the data set $\mathbf{x}$. For the exponentially spaced data set $\mathbf{x}^\tau$, convergence was never attained, not even when the relative change in the parameters is set to one decimal place.

[8] All computations were done in Ox, see Doornik [28].

MLE(BFGS)-Iter.	l_θ	a	b	ρ
Initial	-321.49	0.20000	1.0000	0.20000
5	449.82	0.71512	1.2195	0.86662
10	461.64	0.76833	1.6685	0.82926
15	480.24	0.99379	3.3552	0.68273
19	480.24	0.99607	3.3601	0.68259

Table 6.2: BFGS iterations for the MLE estimation corresponding to the simulated Poisson-gamma model ($a = 1, b = 3$ and $\rho = 0.7$) based on the uniformly simulated data ($\mathbf{x}$).

MCEM-Iter.	MC-sim (m)	$\hat{Q}$	l_θ	a	b	ρ
Model			478.72	1.0	3.0	0.7
1	30	-35.918	281.74	0.99767	3.3630	0.12458
2	15	-62.689	294.19	1.0081	3.3982	0.14979
3	8	-122.25	309.94	1.0076	3.3966	0.18435
4	5	-148.64	322.86	1.0070	3.3946	0.21350
5	5	-190.99	339.73	1.0066	3.3934	0.25236
10	5	–345.00	398.47	1.0111	3.4088	0.39295
20	5	-525.91	454.29	0.99835	3.3666	0.54879
30	5	-634.71	474.07	1.0055	3.3913	0.62272
40	5	-706.49	479.18	0.99591	3.3593	0.66021
50	5	-695.56	479.21	1.0154	3.4251	0.65675
60	5	-724.07	480.03	0.99851	3.3682	0.67218
70	5	-743.83	480.21	1.0030	3.3833	0.67803
80	5	-721.39	479.96	1.0155	3.4255	0.66841
90	5	-722.50	479.90	1.0063	3.3945	0.66816
100	5	-774.23	480.23	0.99182	3.3458	0.68584

Table 6.3: MCEM iterations for the simulated Poisson-gamma model ($a = 1, b = 3$ and $\rho = 0.7$) based on the uniformly simulated data ($\mathbf{x}$). For the results in this table the MC simulations required within each iteration where reduced by half in each iteration with a minimum of 5 simulations.

MCEM-Iter.	MC-sim (m)	$\hat{Q}$	l_θ	a	b	ρ
Model			374.48	1.0	3.0	0.7
1	30	-120.80	142.73	0.95280	2.6325	0.063404
2	15	-32.749	153.78	1.0729	3.1216	0.053279
3	8	-27.059	157.77	1.0895	3.1878	0.060540
4	5	-30.732	159.52	1.0926	3.1982	0.064691
5	5	-37.874	163.66	1.0793	3.1529	0.078280
10	5	-79.441	185.96	1.0781	3.1455	0.14994
20	5	-135.59	219.68	1.0837	3.1561	0.26650
30	5	-216.14	250.83	1.0487	3.0281	0.38240
40	5	-249.71	263.20	1.0465	3.0232	0.42071
50	5	-266.83	267.86	1.0321	2.9609	0.44180
60	5	-279.02	269.99	1.0256	2.9453	0.44989
70	5	-272.45	269.77	1.0314	2.9563	0.44792
80	5	-274.45	270.91	1.0320	2.9621	0.45049
90	5	-269.96	269.86	1.0360	2.9763	0.44575
100	5	-263.01	269.29	1.0551	3.0484	0.43593

Table 6.4: MCEM iterations for the simulated Poisson-gamma model ($a = 1, b = 3$ and $\rho = 0.7$) based on the exponentially spaced data ($\mathbf{x}^\tau$). For the results in this table the MC simulations required within each iteration where reduced by half in each iteration with a minimum of 5 simulations.

CHAPTER 7

LATENT STRUCTURE BASED MODELS: NONPARAMETRIC APPROACH

For the Gibbs type construction of stationary models, a parametric form for the required conditional distributions $F_{Y|X}$ and $F_{X|Y}$ has to be completely specified. In this chapter we relax the parametric specification of such conditional distributions by using some ideas from the nonparametric Bayesian literature. Section 7.1 introduces a method to construct stationary autoregressive models with arbitrary stationary distribution. The latter method uses certain random probability measures. We have explored two choices for such measures, one based on Gaussian processes and other based on Pólya trees. These are given in Section 7.3 and Section 7.4 respectively.

7.1 NONPARAMETRIC TIME SERIES MODEL

The construction of stationary models presented in Chapter 3 relied on the prior knowledge of the parametric form for the conditional distribution $F_{Y|X}$ (or equivalently $F_{X|Y}$ and Q_Y). Some of the aspects considered when choosing this parametric family were discussed in Section 3.5 and played the driving mechanism for most of the models introduced so far. For example, in Chapter 4, the choice of the conditional $F_{Y|X}$ was done in a way such that the resulting models were of the ARCH-type.

Fixing a specific parametric form for the conditional $F_{Y|X}$ can be restrictive in some modelling contexts. For example, if we aim to construct a stationary model for which the only knowledge (or assumption) about its behavior is its invariant distribution, then

imposing a parametric structure for $F_{Y|X}$ may induce an unrealistic correlation structure for the problem at issue. In order to circumvent this "unrealistic" assumption, a nonparametric approach can be taken.

The independence structure (3.17), underlying to the Gibbs type construction, forces the target process $\{X_t\}_{t=1}^{\infty}$ to be conditionally independent given certain elements of the latent sequence $\{Y_t\}_{t=1}^{\infty}$. This in turn leads to reversible models, implying that the corresponding *fdds* are symmetric. A stronger type of symmetry, that the one induced by reversible sequences, is the one presented by exchangeable sequences. That is, a sequence of random variables for which the corresponding *fdds* remain invariant under permutations. The strength of exchangeable sequences derives from De Finetti's representation theorem that, loosely speaking, establishes that given an infinite exchangeable sequence $\{X_t\}_{t=1}^{\infty}$, there exists a random quantity F such that the X_i are IID given F[1]. We refer to Schervish [93] for proofs of De Finetti's representation theorem.

The above discussion suggests that instead of conditioning $\{X_t\}_{t=1}^{\infty}$ on a particular characteristic of interest, such as any parameter/information represented by the latent variables "Y's", we can condition on $\mathrm{F} \in \mathcal{F}$, where $\mathcal{F}$ denotes the space of all probability measures on $(\mathrm{E}, \mathcal{E})$. This in turn implies that we would need to compute $(\mathrm{E}, \mathcal{E})$-measurable conditional probabilities given certain probability measures which are elements of $\mathcal{F}$. Therefore, by definition of conditional probability, this $(\mathrm{E}, \mathcal{E})$-measurable conditional probabilities must be measurable in some space. Assume that this space is the Borel σ-field $\mathcal{B}_F$ generated by the topology of weak convergence.

In this chapter, we assume that we want to construct stationary autoregressive models (so $\mathcal{T}$ is countable), with state space E being $\mathbb{R}$ or $\mathbb{R}_+$. As before, we denote with Q_X the probability measure induced by the parametric form assumed for the stationary distribution. Here, due to the state space assumptions, Q_X is set to be diffuse. Namely, we will only consider AR models with continuous stationary distributions.

Analogously to the Gibbs type construction, in this case we base the construction on

[1]It is worth noting that, even though conditionally IID implies exchangeability, a sequence of exchangeable random variables is not always conditionally IID. See Schervish [93]

the following joint distribution defined on $E \times \mathcal{F}$

$$\begin{aligned} \mathrm{P}(X_t \in B, F \in A) &= \mathbb{E}_{\mathcal{P}}\{F(B)\,\mathbb{1}(F \in A)\} \\ &= \int_A F(B)\,d\mathcal{P}(F), \end{aligned}$$

where $B \in \mathcal{E}$, $A \in \mathcal{B}_F$ and $\mathcal{P}$ is a measure on $(\mathcal{F}, \mathcal{B}_F)$. Integrating out F, define

$$Q_X(B) = \mathbb{E}_{\mathcal{P}}\{F(B)\} = \int_{\mathcal{F}} F(B)\,d\mathcal{P}(F). \tag{7.1}$$

Hence the "posterior" probability for F given X_t is given by

$$\mathrm{P}(F \in A \mid X_t \in B) = \frac{\mathbb{E}_{\mathcal{P}}\{F(B)\,\mathbb{1}(F \in A)\}}{Q_X(B)}. \tag{7.2}$$

The conditional probability (7.2), plus the assumption that the X's given F are IID F, namely

$$\mathrm{P}(X \in B \mid F) = F(B), \tag{7.3}$$

provides us with a way to construct the following one-step transition probability

$$\begin{aligned} \mathrm{P}(x, B) &= \mathrm{P}(X_t \in B \mid X_{t-1} = x) \\ &= \mathbb{E}\{F(B) \mid X_{t-1} = x\} \\ &= \frac{1}{Q_X(dx)} \int_{\mathcal{F}} F(B)F(dx)\,d\mathcal{P}(F). \end{aligned} \tag{7.4}$$

The expectation in the right hand side of expression (7.4) is taken with respect to the measure (7.2). Within this nonparametric framework the conditional probability (7.2) plays the role of $F_{Y|X}$ whereas (7.3) plays the role of $F_{X|Y}$.

The construction of the above one-step transition probability is equivalent to finding predictive probabilities based on one observation within the Bayesian nonparametric literature. In the latter context, the random probability measure $\mathcal{P}$, can be seen as a nonparametric prior. An important property of the target process X, also satisfied in this case, is the underlying reversibility. The latter property might be also useful when analyzing predictive distributions in the nonparametric Bayesian framework.

Notice that the structure required for Q_X, equation (7.1), encompass all continuous probability measures defined on E, after a suitable choice for the random measure

$\mathcal{P}$. This will allow us to construct stationary models with a wide choice of invariant measures. The following proposition ensures the stationarity of the model.

Proposition 7.1. If $X_{t-1} \sim Q_X$ and $X_t \mid X_{t-1} = x \sim \mathrm{P}^x$ with $\mathrm{P}^x = \mathbb{E}\{F \mid X_{t-1} = x\}$ then marginally $X_t \sim Q_X$.

Proof. Using the transition probability (7.4),

$$\begin{aligned} \mathrm{P}(X_t \in B) &= \int_E \int_{\mathcal{F}} F(B)\, F(dx)\, d\mathcal{P}(F) \\ &= \int_E \int_{\mathcal{F}} F(B)\, F(dx)\, d\mathcal{P}(F) \end{aligned}$$

where the second equality follows from Fubini's theorem and the fact that F is a probability measure. □

7.2 AR(1) MODELS BASED ON DIRICHLET PROCESSES

Pitt et al. [78] proved that the DAR(1) model, see McDonald and Zucchini [71], can be seen as an AR(1) process constructed using the mechanism described in Section 7.1 when the random probability measure $\mathcal{P}$ is set to be the Dirichlet process.

More precisely, if we denote by $\mathcal{D}_Q^c$ a Dirichlet process driven by the measure $cQ(\cdot)$, where $c > 0$ and Q is a probability measure on $(\mathrm{E}, \mathcal{E})$. Hence a random probability measure distributed accordingly with a Dirichlet process, denoted by $F \sim \mathcal{D}_Q^c$, satisfies

$$\mathbb{E}_{\mathcal{D}_Q^c}[F(B)] = Q(B)$$

for any $B \in \mathcal{E}$. See Ferguson [35]. The parameter $c > 0$ is commonly associated to the variability of the random probability measure F about Q.

In this case, the well-known conjugacy of the Dirichlet process leads to

$$F \mid X \sim \mathcal{D}_{Q^*}^{c+1},$$

where

$$Q^*(B) = \frac{c}{c+1} Q(B) + \frac{1}{c+1} \delta_X(B),$$

and δ_X denotes the point mass at X. See Theorem 1 in Ferguson [35]. Therefore, using the interpretation presented in Section 7.1, we can construct the following transition probability

$$\begin{aligned} \mathrm{P}(x, B) &= \mathbb{E}\{F(B) \mid X_{t-1} = x\} \\ &= \frac{c}{c+1} Q(B) + \frac{1}{c+1} \delta_X(B), \end{aligned} \tag{7.5}$$

which satisfies Proposition 7.1. In particular, if a density function q corresponding to the probability measure Q exists, then the density corresponding to the above transition is given by

$$p(x_{t-1}, x_t) = \frac{c\, q(x_t) + \delta_{x_{t-1}}(x_t)}{c+1}. \tag{7.6}$$

Consequently, we have

$$\mathrm{P}(X_t = X_{t-1}) = \frac{1}{c+1}.$$

Model (7.6) is known as the DAR(1) model. Some other discrete time series can arise by changing the assumptions for the measure $\mathcal{P}$. For example, by setting $\mathcal{P}$ to be a random probability measure constructed by normalizing increasing additive processes, see James et al. [46] for more on these random measures.

Therefore, the issue here is which $\mathcal{P}$ to use in order to be able to manipulate the one-step transition probability (7.4). A constraint we put on the choice of the random probability measure $\mathcal{P}$ is that it should put probability one to the set of all continuous or absolutely continuous distributions. In other words, we look for models where

$$\mathrm{P}(X_t = x \mid X_{t-1} = x) = 0.$$

The above constraint rules out random measures that put positive mass on discrete distributions, such as the Dirichlet process[2] and those based on normalized increasing additive processes. In the following two sections we study different possibilities for the choice of the random probability measure $\mathcal{P}$.

[2]One way to tackle the discreteness of Dirichlet processes is via mixtures, either via randomizing the "baseline measure", Antoniak [4] or by using a kernel mixture as in Lo [66].

7.3 AR(1) MODELS BASED ON GENERALIZED LOG-GAUSSIAN PROCESSES

A way to construct random measures $\mathcal{P}$ on the space $(\mathcal{F}, \mathcal{B}_F)$ can be done by suitably normalizing certain stochastic process $\{W_t, t \geq 0\}$. The main contextual difference is that instead of considering realizations on the "time" space we consider realizations on the state space, that is we set $\mathcal{T}$ to be equal to E, the support of the underlying probability measures. However, thinking of a realization of a process $\{W(t); t \in \mathcal{T}\}$ as a probability measure belonging to the space $\mathcal{F}$, imposes some constraints on such realizations. For instance, the clearest one must be that such realizations must characterize probability measures, for example via cumulative distribution functions, survival functions, densities, etc.

In this section we use a method based on Gaussian processes studied by Lenk [62] based on ideas of Leonard [63] and Thorburn [97]. In order to introduce this method first let us consider the family of absolutely continuous distributions with respect to λ, where λ a finite measure on the Borel space $(\mathrm{E}, \mathcal{E})$. Denote the corresponding densities as f. Therefore, using the notation in Definition 2.1, we are interested in a stochastic process on a certain probability space $(\Omega, \mathcal{A}, \mathrm{P})$, where $\mathcal{T} = \mathrm{E}$. In other words, a density f is modelled by a stochastic process $f = \{f(x, \omega); x \in \mathrm{E}, \omega \in \Omega\}$ with state space $(\mathbb{R}_+, \mathcal{B}(\mathbb{R}_+))$. For a fixed ω, a "sample path", $f(\cdot)$ stands for the density corresponding to a probability measure with support E. With this, we denote $(\mathbb{R}_+^E, \mathcal{B}(\mathbb{R}_+)^E)$ as the measurable space of functions from E to $\mathbb{R}_+$ and $(\mathcal{F}_d, \mathcal{B}_d, \mathcal{P}_d)$ as the probability space of densities with support E. Here,

$$\mathcal{F}_d = \{f; f \text{ is a density with support E}\}$$

and $\mathcal{B}_d$ is the restriction of $\mathcal{B}(\mathbb{R}_+)^E$ to $\mathcal{F}_d$.

Working under the framework described above, we can write the transition density corresponding to the probability measure (7.4) and stationary density q_X, as follows

$$p(x_{t-1}, x_t) = \frac{1}{q_X(x_{t-1})} \int_{\mathcal{F}_d} f(x_t)\, f(x_{t-1})\, d\mathcal{P}_d(f), \tag{7.7}$$

Where $q_X = \mathbb{E}_{\mathcal{P}_d}(f)$.

The logistic normal distribution

The approach taking here resembles the one described in Lenk [62] within a Bayesian nonparametric framework. Let Z be a Gaussian process with mean function μ and covariance function σ. Assume that μ is bounded on E, σ and bounded on E^2. Furthermore, suppose that there exist positive constants C, β, ϵ and a non-negative integer r such that

$$\mathbb{E}|Z(x) - Z(y)|^{\beta} \leq C|x - y|^{1+r+\epsilon}.$$

Then Kolmogorov's criterion is satisfied, and an r times continuously differentiable modification of Z exists. Consider the following normalization of log-Gaussian processes

$$f(x) = \frac{W(x)}{\int_E W(s)\, d\lambda(s)}, \tag{7.8}$$

where $\lambda(\mathrm{E}) < \infty$ and $W = \exp\{Z\}$. Its distribution Λ is denoted by $\mathrm{LN}_E(\mu, \sigma)$. Provided the existence of such process, f has sample paths that are densities, and its support contains the densities on E with respect to λ.

The joint moments of W are given by

$$\begin{aligned} \mathrm{M}(\mathbf{x}) &= \mathbb{E}\left[\prod_{k=1}^{K} W(x_k)\right] \\ &= \exp\left\{\left(\sum_{k=1}^{K} \mu(x_k) + \frac{\sigma(x_k, x_k)}{2}\right) + \sum_{i<j}^{K} \sigma(x_i, x_j)\right\}, \end{aligned} \tag{7.9}$$

where $\mathbf{x} \in \mathbb{R}^K$. The random density f uses the path-integral of W, for which the positive moments are given by

$$\begin{aligned} \mathrm{C}(K, \mu) &= \mathbb{E}\left[\left(\int_E W d\lambda\right)^K\right] \\ &= \int_{E^K} \mathbb{E}\left[\prod_{k=1}^{K} W(s_k)\right] d\lambda(\mathbf{s}) \\ &= \int_{E^K} \exp\left\{\sum_{i<j}^{K} \sigma(s_i, s_j)\right\} \prod_{k=1}^{K} W_0(s_k)\, d\lambda(\mathbf{s}), \end{aligned} \tag{7.10}$$

where $d\lambda(\mathbf{s}) = d\lambda(s_1)d\lambda(s_2)\cdots d\lambda(s_K)$ and

$$W_0(s) = \mathbb{E}[W(s)] = \exp\{\mu(s) + \sigma(s,s)/2\}.$$

The first equality in (7.10) follows by an application of Fubini's theorem.

It can be proved that the use of process (7.8) as nonparametric prior leads to non-conjugate posteriors (7.2). In theory, this issue does not affect our construction of stationary models, however in order to get the required transition probabilities, the computations are simplified when using conjugate random measures. In order to circumvent this issue Lenk [62] considered a generalization of the random density (7.8) characterized by the following distribution

$$\Lambda_\xi(A) = \frac{\int_A \left\{\int_E W(s)\, d\lambda(s)\right\}^\xi d\Lambda(W)}{\mathrm{C}(\xi,\mu)}, \tag{7.11}$$

where $\Lambda \sim W \sim \mathrm{LN}_E(\mu,\sigma)$. The generalized distribution (7.11) is indicated by $\mathrm{LN}_E(\mu,\sigma,\xi)$.

In [62], the ***logistic normal process***, denoted by $\mathrm{LNS}_E(\mu,\sigma,\xi)$, is defined as the random density on $(\mathcal{F}_d,\mathcal{B}_d)$ defined by (7.8) with $W \sim \mathrm{LN}_E(\mu,\sigma,\xi)$.

Theorem 7.1. (Corollary 3 in [62]) Let $f \sim \mathrm{LNS}_E(\mu,\sigma,\xi)$ and $\mathbf{x} \in \mathrm{E}^K$, then

$$\mathbb{E}\left[\prod_{i=1}^K f(x_i)\right] = \mathrm{M}(\mathbf{x})\,\frac{\mathrm{C}(\xi-K,\mu^*)}{\mathrm{C}(\xi,\mu)},$$

where

$$\mu^*(s) = \mu(s) + \sum_{i=1}^K \sigma(s,x_i).$$

Proof. For the proof we refer to Lenk [62]. □

Hence, modelling the random densities as $f \sim \mathrm{LNS}_E(\mu,\sigma,\xi)$ the transition density (7.7) is given by

$$\begin{aligned} p(x_{t-1},x_t) &= \frac{\mathbb{E}\left[f(x_t)f(x_{t-1})\right]}{q_X(x_{t-1})} \\ &= \frac{1}{q_X(x_{t-1})}\,\mathrm{M}((x_t,x_{t-1})')\,\frac{\mathrm{C}(\xi-2,\mu^\circ)}{\mathrm{C}(\xi,\mu)}, \end{aligned} \tag{7.12}$$

where $q_X(x) = \mathbb{E}[f(x)]$ and

$$\mu^{\diamond}(s) = \mu(s) + \sigma(s, x_t) + \sigma(s, x_{t-1}).$$

Using Theorem 7.1, the stationary density q_X can be represented as follows

$$q_X(x) = \mathbb{E}[f(x)] = \exp\{\mu(x) + \sigma(x,x)/2\} \frac{\mathrm{C}(\xi - 1, \mu^x)}{\mathrm{C}(\xi, \mu)},$$

where $\mu^x(s) = \mu(s) + \sigma(s, x)$.

In order to ease the computations, let us assume that the random densities are modelled by $f \sim \mathrm{LNS}_E(\mu, \sigma, 1)$ then

$$q_X(x) \propto \exp\{\mu(x) + \sigma(x,x)/2\}. \tag{7.13}$$

Therefore, if we want to construct a stationary process with invariant density being q_X, then the required random measure for the Gibbs type construction via conditionals (7.2) and (7.3) can be chosen by fixing a valid covariance function σ, and then from (7.13), by taking

$$\mu_0(x) = \ln(q_X(x)b) - \frac{\sigma(x,x)}{2}, \tag{7.14}$$

where $b > 0$.

Proposition 7.2. Let us assume that q_X and σ are known and that $f \sim \mathrm{LNS}_X(\mu_0, \sigma, 1)$ then

$$p(x_{t-1}, x_t) \propto q_X(x_t) \exp\{\sigma(x_t, x_{t-1})\} \mathrm{C}(-1, \mu^{\diamond}). \tag{7.15}$$

Proof. The proof is given by direct substitution of (7.14) in (7.12) with M given by expression (7.9). □

Hence, regardless of the choice of q_X we can construct an AR(1) model (7.15) with invariant measure which density is given by q_X.

As pointed out in [62], the dependence structure of the process $\{X_t\}_{t=1}^{\infty}$ is driven by the covariance function σ chosen for the required log-Gaussian process. If $\sigma \to 0$

then the model approach to the independent case, meaning that not lag dependence is presented. Hence, in order to construct a transition density that leads to a stationary process with given invariant distribution a valid covariance function has to be specified. This is precisely where the freedom in the choice of the random measure lies.

In this approach, the latent structure required in the Gibbs type construction, has a nonparametric form. At first, this suggests that results such as those obtained in Section 4.10 (concerned with the parameter estimation of a certain target process) might be more complicated, in the sense that we need to sample from a nonparametric posterior. For some well known random probability measures, such as the Dirichlet process, techniques circumventing the latter issue exists within the Bayesian nonparametric literature. Due to the "Gaussianity" of the approach presented in this section, the analysis of the latent structure (or nonparametric posterior) may be relatively simple. However, we leave that issue for future research. Here, we will concentrate in the transition density directly.

In Proposition 7.2 we could have used $\mathrm{LNS}_X(\mu_0, \sigma, 2)$ instead of $\mathrm{LNS}_X(\mu_0, \sigma, 1)$, this would ease the computations since the negative moments $\mathrm{C}(-1, \mu^\diamond)$ are no longer required. However, the interpretability of the method is not very clear in this case, since one process ($\xi = 1$) is used to constrain the mean function μ_0 of the underlying Gaussian process and another process ($\xi = 2$) is used to get the transition.

The problem with negative moments is that they are not as tractable as the positive ones. In fact, if we have a positive random variable Y with finite moments of any other

$$\mathbb{E}[Y^{-r}] = \frac{1}{(r-1)!} \int_0^\infty u^{r-1}\, \mathbb{E}[e^{-uY}]\, du,$$

where

$$\mathbb{E}[e^{-uY}] = \sum_{k=0}^{\infty} (-1)^k \frac{u^k}{k!}\, \mathbb{E}[Y^k].$$

Hence taking $Y = \int W d\lambda$ we get

$$\mathrm{C}(-r, \mu^*) = \frac{1}{(r-1)!} \int_0^\infty \sum_{k=0}^{\infty} (-1)^k \frac{u^{k+r-1}}{k!}\, \mathrm{C}(k, \mu^*)\, du.$$

Lenk [62] proposed two methods to compute the negative moments. The first, which

is given after two applications of the mean value theorem, gives the following

$$\mathrm{C}(-r, \mu^*) \propto \exp\{-r\sigma(x_t, \theta)\},$$

where θ depends on x_t, x_{t-1}, q_X and σ is approximated by a cubic spline. The second, which is based in a Monte Carlo approximation, results in

$$\mathrm{C}(-r, \mu^*) \propto \exp\{-r\sigma(x_t, +)\},$$

where

$$\sigma(x, +) = \int_E \sigma(x, s)\, d\lambda(s)/\lambda(\mathrm{E}).$$

For any of the two methods above described, the transition density (7.15) can approximated by

$$p(x_{t-1}, x_t) \propto q_X(x_t) \exp\{\sigma(x_t, x_{t-1}) - \sigma(x_t, \cdot)\}. \tag{7.16}$$

Example 7.1. Let us assume that we want to construct a stationary model with invariant distribution $Q_X = \mathrm{N}(0, 1)$. Following the approach presented in this section, we need to specify a valid covariance function such that the resulting Gaussian process has smooth paths. Smoothness of the paths of a Gaussian process can be ensured by smoothness of the underlying covariance function. In particular, when dealing with stationary Gaussian process (namely $\sigma(x, y) = \sigma^*(|x - y|)$) it suffices to check that the stationary covariance function $\sigma(x)$ is differentiable at zero. For the sake of exposition we use the following two choices

$$\text{Gaussian:} \qquad \sigma(x) = \alpha e^{\beta x^2} \tag{7.17}$$

$$\text{Bilateral exponential:} \qquad \sigma(x) = \frac{\alpha}{1 + \beta x^2}, \tag{7.18}$$

where $\alpha, \beta > 0$. The parameter β gives the strength in which the lag-dependence is taken into account, in general when $\beta \to 0$, the model tends to non lag-dependence. Figure 7.1 shows the density estimates and autocorrelation functions corresponding to a random sample of size 1000 from the target process with transition densities (7.16). The simulations were done using the inverse CDF method. The required normalization was done with numerical integration. ∘

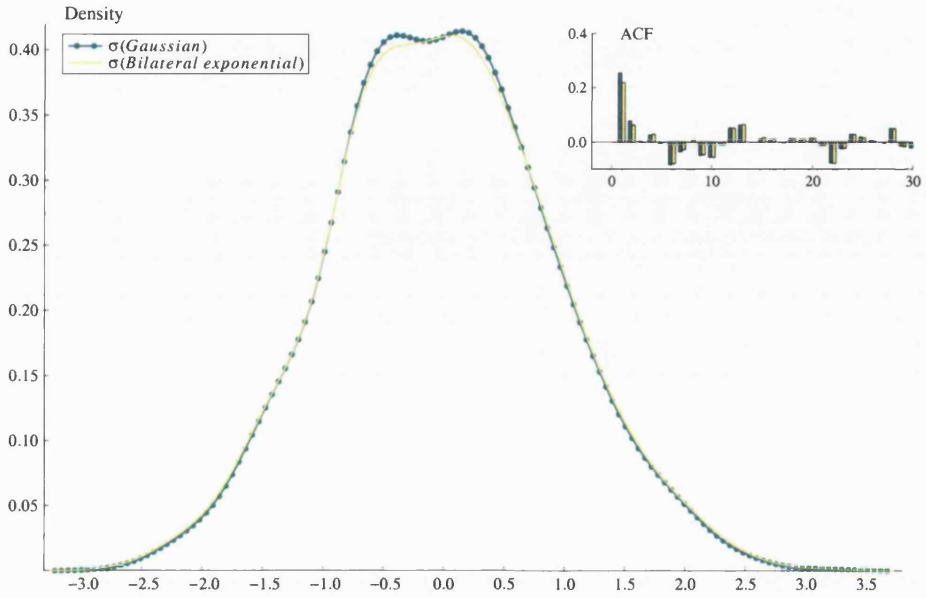


Figure 7.1: Density estimates and ACF's corresponding to 1000 simulations from stationary processes with transition densities (7.16). The simulations were done for the two different covariance functions given by (7.17) and (7.18) with $\alpha = 1, \beta = 1$.

Using the normalized transition densities, estimation of the parameters involved in the covariance functions (in general, parameters in q_X can also be included) can be done using the BFGS algorithm by directly using the transition density. Alternatively, Monte Carlo methods such as the Metropolis-Hastings algorithm can improved the parameter estimation by avoiding the numerical normalization required in transition probability given by equation (7.16).

7.4 AR(1) MODELS BASED ON PÓLYA TREES

In this section we use a different choice for the random probability measure $\mathcal{P}$. Let $\Pi = \{B_\epsilon\}$ be a binary partition of the state space $(\mathrm{E}, \mathcal{E})$ where $\epsilon = \epsilon_1 \cdots \epsilon_m$, $\epsilon_j \in \{0, 1\}$. The subindex ϵ allocates the set B_ϵ in the tree while keeping the branch information. The partition mechanism is given in the following way: In the mth level partition[3] B_ϵ splits into $(B_{\epsilon 0}, B_{\epsilon 1})$, $B_{\epsilon 0}$ into $(B_{\epsilon 00}, B_{\epsilon 01})$ and so forth until infinity. Random mass is allocated to the sets via independent beta random variables $Y_{\epsilon 0} \sim \mathrm{Be}(\alpha_{\epsilon 0}, \alpha_{\epsilon 0})$, $Y_{\epsilon 1} = 1 - Y_{\epsilon 0}$ for non-negative numbers $\alpha_{\epsilon 0}$ and $\alpha_{\epsilon 0}$. Then, at a given level m the mass

[3]Remember that by definition of partition $B_{\epsilon 0} \cap B_{\epsilon 1} = \emptyset$ and $B_{\epsilon 0} \cup B_{\epsilon 1} = B_\epsilon$.

allocated to a particular set is given by

$$\mathrm{F}(B_\epsilon) = \left(\prod_{j=1;\epsilon_j=0}^{m} Y_{\epsilon_1 \cdots \epsilon_{j-1}0} \right) \left(\prod_{j=1;\epsilon_j=1}^{m} 1 - Y_{\epsilon_1 \cdots \epsilon_{j-1}0} \right),$$

where $\epsilon = \epsilon_1 \cdots \epsilon_m$. See Lavine [57], Muliere and Walker [76] and the references therein for a more exhaustive treatment of Pólya trees. In theory the number of levels required is infinity, however an approximation is commonly used by terminating the process at a finite level m. The standard notation for random probability measures with Pólya tree distribution is $F \sim \mathrm{PT}(\Pi, \mathscr{A})$. The Dirichlet process arises when $\alpha_{\epsilon 0} + \alpha_{\epsilon 1} = \alpha_\epsilon$ for all ϵ, see Ferguson [36].

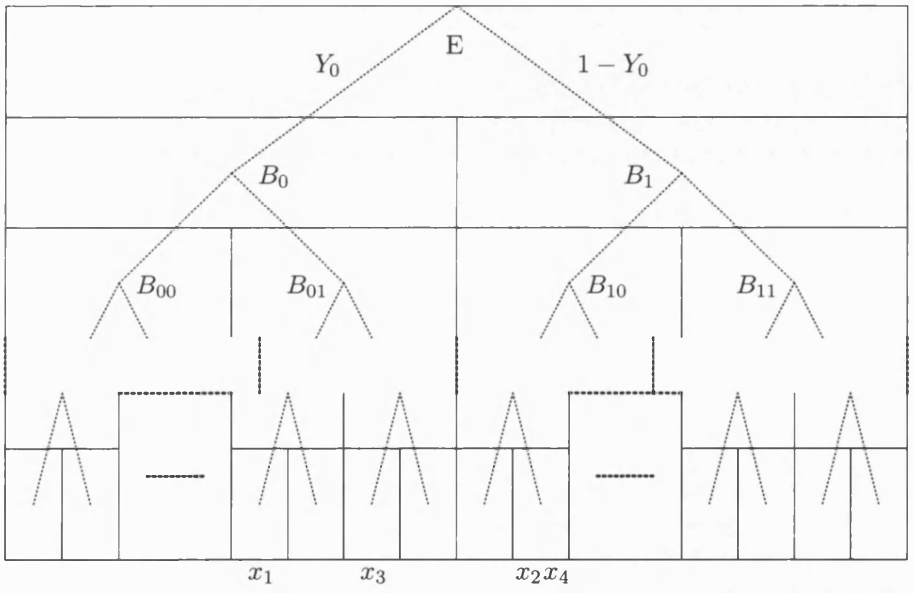


Figure 7.2: Pólya tree distribution on E. The values x_1 and x_2 only share the partition set E ($k = 1$), whereas x_1 and x_3 share E, B_0 and B_{01} ($k = 3$). For x_2 and x_4, $k = m$.

It is possible to center the process in a specific distribution Q_X, that is $\mathbb{E}[\mathrm{F}] = Q_X$, in this case our known invariant distribution for the AR(1) model. The are, at least, two methods to implement the latter centering procedure; by taking $\alpha_{\epsilon 0} = c_\epsilon Q_X(B_{\epsilon 0})$ and $\alpha_{\epsilon 1} = c_\epsilon Q_X(B_{\epsilon 1})$, for some $c_\epsilon > 0$, or alternatively by taking the partitions to coincide with the percentiles of Q_X and restricting $\alpha_{\epsilon 0} = \alpha_{\epsilon 1}$ for all ϵ. For example, if $\mathrm{E} = \mathbb{R}$ the first partition is chosen to be $B_0 = (-\infty, Q_X^{-1}(\frac{1}{2}))$ and $B_1 = [Q_X^{-1}(\frac{1}{2}), \infty)$. In general

the partitions at level m for $\xi = 1, \ldots, 2^m$ are given by

$$B_{m\xi} = \left[Q_X^{-1}\left(\frac{\xi - 1}{2^m}\right), Q_X^{-1}\left(\frac{\xi}{2^m}\right)\right), \tag{7.19}$$

with $Q_X^{-1}(0) = -\infty$ and $Q_X^{-1}(1) = \infty$. See Ferguson [36], Mauldin et al. [70], Lavine [57] and Walker et al. [102].

In this case the nonparametric latent structure is given through the posterior (7.2) which in this case is easily found to be $\mathrm{PT}(\Pi, \mathscr{A} \mid X_t)$ where

$$\mathscr{A} \mid X_t = \begin{cases} \alpha_\epsilon + 1 & \text{if } X_t \in B_\epsilon \\ \alpha_\epsilon & \text{otherwise.} \end{cases} \tag{7.20}$$

As we have mentioned before, a desirable property of the chosen random probability measure $\mathcal{P}$ is that it leads to transition probabilities that satisfy $\mathrm{P}(X_t = x \mid X_{t-1} = x) = 0$ for all x. For Pólya tree distributions the probability of coincidence is given by

$$\mathrm{P}(X_t = x \mid X_{t-1} = x) = \prod_{k=1}^{\infty} \frac{\alpha_{\epsilon_1 \cdots \epsilon_k} + 1}{\alpha_{\epsilon_1 \cdots \epsilon_{k-1} 0} + \alpha_{\epsilon_1 \cdots \epsilon_{k-1} 1} + 1}, \tag{7.21}$$

where $x \in B_{\epsilon_1 \cdots \epsilon_k}$ for all $k = 1, 2, \ldots$ This tends to zero provided that the α_ϵ does not decrease too rapidly.

Under the conditions of Dirichlet process, this probability is

$$\frac{1}{\alpha_0 + \alpha_1 + 1} > 0$$

and $\alpha_0 + \alpha_1 = c$. See Lavine [57].

Here we take all α_ϵ to be constant and fixed at the value $c > 0$. This ensures $\mathrm{P}(X_t = x \mid X_{t-1} = x) = 0$ for all x, since

$$\lim_{n \to \infty} \left(\frac{c+1}{2c+1}\right)^n = 0.$$

We center the Pólya tree in Q_X by chosen the partitions to coincide with the percentiles of Q_X.

Given the stationary distribution Q_X, if the random measure in (7.4) is modelled by $F \sim \mathrm{PT}(\Pi, \mathscr{A})$ and all α_ϵ are fixed to be $c > 0$, then the transition probability

corresponding to a stationary AR(1) model is given by

$$\mathrm{P}(X_t \in \mathrm{B}_\epsilon \mid X_{t-1} = x) = \begin{cases} \left(\frac{c+1}{2c+1}\right)^{k_t} \frac{c}{2c+1} \left(\frac{1}{2}\right)^{m-k_t-1} & 0 \leq k_t \leq m \\ \left(\frac{c+1}{2c+1}\right)^{k_t} & k_t = m, \end{cases} \tag{7.22}$$

where $\epsilon = \epsilon_1 \cdots \epsilon_m$ and k_t denotes the number of levels in which both X_{t-1} and X_t share the same partition set. Figure 7.2 illustrates how to obtain k.

The distribution of k is seen to be geometric; that is,

$$p(k) = \rho(1-\rho)^k, k = 0, 1, \ldots$$

and $\rho = c/(2c+1)$. Consequently, $\mathbb{E}[k] = 1 + 1/c$, providing clear interpretation for c.

This choice ensures that a random distribution chosen from such a Pólya tree is absolutely continuous and that $\mathrm{P}(X_t = X_{t-1}) = 0$. The geometric distribution of k can be seen as a consequence of the stationarity assumption.

A similar approach of constructing autoregressive process was introduced by Sarno [89]. She constructed the transition by taking the number of levels with shared partitions to be geometric but with the α_ϵ, required at each level, being $\alpha_\epsilon = m^2$. This choice also ensures the probability of coincidence to be zero. However, taking the α_ϵ to be constant suffices for the latter condition. In general, the model of Sarno [89] is not stationary but a modification to it, which does not compromise the model in any way, guarantees stationarity.

Another advantage of the specification $\alpha_\epsilon = c$ is the simple form of the likelihood estimator for c. Let the number of levels m to approach to infinity in the transition probability (7.22), the score for c corresponding to a sample $\mathbf{x} = (x_1, x_2, \ldots, x_T)$ is given by

$$\frac{\partial \log L_{\mathbf{x}}(c)}{\partial c} = \frac{T-1}{c(2c+1)} - \frac{1}{(c+1)(2c+1)} \sum_t k_t.$$

By equating the above quantity to zero and solving for c we get the following MLE

$$\hat{c} = \frac{1}{\overline{k} - 1}, \tag{7.23}$$

where $\overline{k}$ denotes the mean of the number of levels shared for consecutive observations.

In conclusion, in order to construct a stationary AR(1) model with Q_X invariant distribution via Pólya trees, we fix the partitions to match the percentiles of Q_X as described in (7.19). Therefore the transition mechanism driving the underlying stationary model is approximated by (7.22).

Notice that the estimator $\hat{c}$ gives the dependence through the sample. A natural question to ask is how this estimator changes as the correlation in the sample varies. In the following example we use simulations to depict the latter relation.

Example 7.2. Let us consider the Gaussian AR(1) model given by

$$Y_t = \rho Y_{t-1} + \sqrt{1 - \rho^2}\epsilon_t, \tag{7.24}$$

where $0 < \rho < 1$ and $\epsilon_t \sim \mathrm{N}(0, 1)$ independently for each t. It is easy to verify that the above model is stationary with $\mathrm{Corr}(Y_t, Y_{t-1}) = \rho$.

In order to illustrate the dependence on the correlation ρ of a given data, of the parameter c, we have simulated 10000 observations of the autoregressive model (7.24) for different values of ρ. For the resulting simulations we fitted a stationary model with invariant distribution $Q_X = \mathrm{N}(0, 1)$ and transition probability (7.22). In other words, given a ρ we simulate from model (7.24) and compute $\hat{c} = 1/(\overline{k} - 1)$. Figure 7.3 shows the results. In general we note that the estimator $\hat{c}$ decreases as the correlation increases. This implies that more levels are shared between consecutive observations (in mean) when the correlation is higher. Analogously when ρ approaches to zero $\hat{c}$ approaches to one, in which case $\mathbb{E}[k] = 1$, meaning that consecutive observations only share one level (in average) when the data are not correlated. ∘

The method presented in this section gives a way to construct stationary AR(1) models with relatively any specification of the invariant distribution. Depending on the problem at issue, one random probability measure might be better than other. Here, we

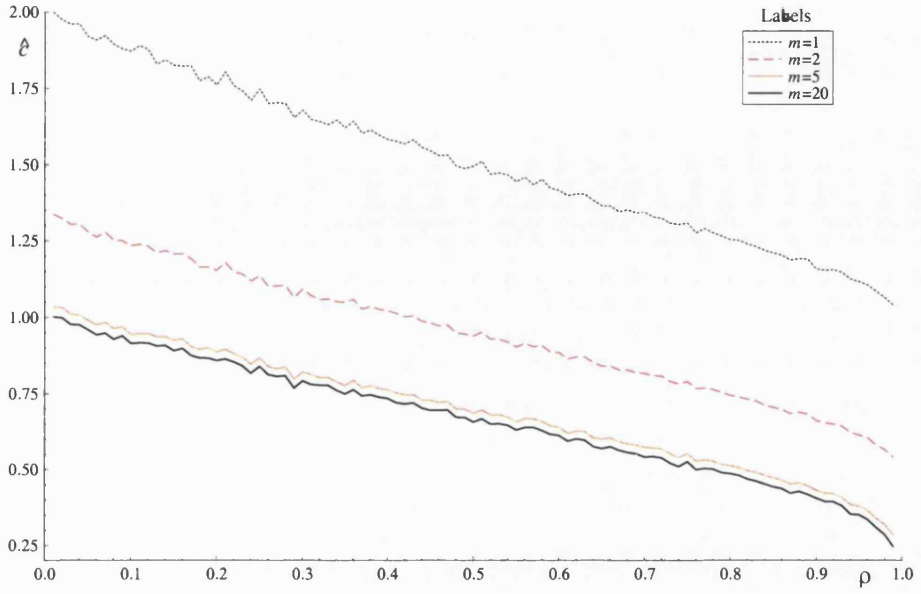


Figure 7.3: Estimator $\hat{c}$ as the correlation ρ varies. The estimator is based on simulated data over for $\rho = 0.01, 0.02, \ldots, 0.99$, see Example 7.2.

have only explored two choices, however some other random probability measures might be used, of course this might or not lead to more tractable transition densities. The main advantage in the nonparametric approach at issue against the parametric specifications studied in previous chapters is the freedom on the underlying latent structure, namely posterior probability. The cost of this freedom comes with the tractability of such quantity, whereas in the parametric case we easily simulated from it, in the nonparametric case it is not always straightforward.

CHAPTER 8

CONCLUSIONS AND FUTURE WORK

The general objective of this thesis was the use of latent structures for constructing and representing stationary models with invariant distributions belonging to a given parametric family. The approaches taken in this thesis to accomplish such objective can be mainly divided into two different routes. The first approach, embracing most of the thesis, Chapters 3, 4, 6 and 7, is based on generalizations of an idea introduced by Pitt et al. [78, 79]. The second approach, presented in Chapter 5, consists of depicting the innovation variable underlying to a self-decomposable random variable which, in particular, eases some estimation and simulation issues of OU-type processes. An important point to emphasize here is that, with the exception of Sections 4.5-4.7, most of the thesis is restricted to time-homogeneous models with the Markov property being satisfied.

We have seen that a generalization of the idea presented in Pitt et al. [78], here termed as the Gibbs sampler type construction, provides with a latent decomposition for the transition probabilities of the model under construction. In particular, we saw that such decomposition eased some estimation and simulation issues. Although other methods to construct stationary models with specific stationary distributions are available in the literature, most of them do not lead to easy estimation techniques. For example, the idea presented in Yacine [108] also studied by Rydberg [88] and generalized by Bibby and Sørensen[15], is devoted to construct the drift and diffusion coefficients of a diffusion process such that a certain specific stationary distribution holds valid. This approach requires more sophisticated estimation techniques; see Bibby and Sørensen [14] and the references therein. On the other hand, the generality encompassed by the

models presented in [15] allows their application in a wide range of applications.

In Pitt et al. [78, 79], the construction was presented only for discrete-time models, in particular a stationary ARCH(1)-type model with Student-t errors is presented in [79]. In the same lines, Chapter 4 generalized the construction of this ARCH-type model in two directions, first to allow for a more general family of innovation distributions and second to allow for higher lag dependence. The outcome of such generalizations led to a stationary ARCH(p)-type model with innovations being generalized hyperbolic distributed. Although, from a different perspective these models were introduced before by Barndorff-Nielsen [9], the latter approach does not ensure stationarity of the model and therefore some estimation results, as the one presented by Theorem 4.2, are not available. Within the econometric framework presented in Section 4.11, the latent structure might be better regarded as a hidden structure since it has the interpretation of being the volatility level given the returns. With this interpretation the latent decomposition may have some applications when analyzing implied volatilities, in which case, calibration methods rather than estimation techniques might be applied to fit the model. We will further look for this applications as well as multivariate versions of stationary ARCH(p) models constructed via the Gibbs sampler type construction.

As we saw before, once the invariant distribution has been fixed, the Gibbs sampler type construction further requires the specification of the parametric family to which the conditional distribution $F_{Y|X}$ belongs (or alternatively the forms for Q_Y and $F_{X|Y}$). Notice that the choice of such parametric family sets the dynamics of the model at issue, that is, different choices of this family lead to different models with the same invariant distribution. In the discrete-time setting, Chapter 7 provides with a way to construct stationary Markov models by relaxing the specification of such families. In general, we allowed such specification to take a nonparametric structure, in which case the specification of suitable probability random measures was required. In particular, we explored two different options for such measures, namely the one based on generalized log-Gaussian processes and the one based on Pólya trees. In both cases a great generality can be accomplish by being able to depict the form for the transition probabilities. However, such generality comes with the price of having nonparametric latent structures and therefore estimation methods via latent decompositions turn out to be more expensive. On the other hand, since transition densities are tractable the

MLE method can be applied.

Following the same ideas presented in Chapter 7, a higher lag-dependance construction seems to be possible, at least, when using Pólya trees. Hence, as future research, we will consider such generalization. When using the generalized log-Gaussian processes, the negative moment required to specify the transition probability in Proposition 7.2 seems to be explicitly available for specific choices of the covariance functions. In particular, when the resulting specification of the covariance function leads to a certain type of diffusion processes then Feynman-Kac formula or extensions to it may help to find the required moment.

In Chapter 3, the Gibbs type construction is also valid for continuous-time processes provided the Chapman-Kolmogorov equations are satisfied. In Chapter 6, we have presented the construction of some continuous-time models that can be identified, to some extent, as diffusion processes. Here the main advantage lies in that, with the Gibbs sampler type construction, we are able to depict the underlying transition density through a latent structure. Which at the same time enabled us to use estimation methods such as the EM algorithm. An important point to underline here is that in all cases considered in Chapter 6 an analytic expression for the transition density was available. However, even in cases where the integration (or summation) required to get the transition density (expression (6.1)) is not possible analytically, the same latent decomposition is available and therefore estimation methods can be applied. In this line we look for continuous-time models for which the transition density is not analytically available whereas the latent structure is. In particular, we look for short interest models without transition densities having the latent decomposition. This would make the parameter estimation of such models easier.

The latent structure representation is not exclusive of stationary processes as we have seen in Chapter 6, where the Poisson-gamma models may lead to squared Bessel process. If a general representation is available, for example for a wide family of diffusion processes, then such representation might be used not only in estimation issues but also in the study and analysis of more general models.

In a different modelling context, another approach to construct stationary models with given invariant distributions (both in discrete and continuous time) is given through

self- decomposable distributions. See, for example, Hart [42], Barndorff-Nielsen et al. [11] and Barndorff-Nielsen and Shephard [13]. As we mentioned before, the difficulty with these approaches is to depict the innovation variable underlying to a SD random variable. Most of the literature available to depict such innovation term is confined to inverse tail Lévy measure based techniques. Chapter 5 provides with a way of understanding the distribution of such innovation term or at least an approximation to it. In particular, a latent representation for such distribution is given. With the approach presented here, inverse tail Lévy measure based methods can be related to simple inverse CDF simulation methods. Furthermore, the simulation schemes for such innovation variables are open for other methods. Barndorff-Nielsen and Shephard [13] have shown the applicability of OU-type processes to volatility modelling. In particular, they emphasized the importance of simulating from the innovation variable corresponding to a SD random variable. With the results presented in Chapter 5 this simulation is improved. Therefore, further developments could be achieved in the volatility modelling presented in [13].

BIBLIOGRAPHY

[1] ABRAHAMSEN, P. A review of Gaussian random fields and correlation functions. Technical Report 917, Norwegian Computing Center., Box 114, Blindern, N-0314 Oslo, Norway, 1997.

[2] ABRAMOWITZ, M., AND STEGUN, I. A., Eds. *Handbook of mathematical functions with formulas, graphs, and mathematical tables.* Dover Publications Inc., New York, 1992. Reprint of the 1972 edition.

[3] ANDERSSON, J. On the normal inverse Gaussian stochastic volatility model. *Journal of Business & Economic Statistics. 19*, 1 (2001), 44–54.

[4] ANTONIAK, C. E. Mixtures of Dirichlet processes with applications to nonparametric problems. *Annals of Statistics 2* (1974), 1152–1174.

[5] ARNOLD, B. C., AND PRESS, S. J. Compatible conditional distributions. *J. Amer. Stat. Assoc. 84* (1989), 152–156.

[6] BAIN, L. J. Moments of non-central t and non-central F-distribution. *Amer. Statist. 23*, 4 (1969), 33–34.

[7] BARNDORFF-NIELSEN, O. E. Exponentially decreasing distributions for the logarithm of particle size. *Proc. Roy. Soc. London A 353* (1977), 401–419.

[8] BARNDORFF-NIELSEN, O. E. *Information and exponential families in statistical theory.* Wiley, Chichester, 1978.

[9] BARNDORFF-NIELSEN, O. E. Normal inverse Gaussian distributions and stochastic volatility modelling. *Scand. J. Statist. 24*, 1 (1997), 1–13.

[10] BARNDORFF-NIELSEN, O. E. Probability and statistics, self-decomposability, finance and turbulence. In *Probability towards 2000 (New York, 1995)*. Springer, New York, 1998, pp. 47–57.

[11] BARNDORFF-NIELSEN, O. E., JENSEN, J., AND SØRENSEN, M. Some stationary processes in discrete and continuous time. *Adv. in Appl. Probab. 30*, 4 (1998), 989–1007.

[12] BARNDORFF-NIELSEN, O. E., AND SHEPHARD, N. Modelling by Lévy processes for financial econometrics. In *Lévy processes. Theory and Applications*, O. E. Barndorff-Nielsen, T. Mikosch, and S. I. Resnick, Eds. Birkhauser, 2001, pp. 283–318.

[13] BARNDORFF-NIELSEN, O. E., AND SHEPHARD, N. Non-Gaussian Ornstein-Uhlenbeck-based models and some of their uses in financial economics (with discussion). *J. R. Statist. Soc. B. 63* (2001), 167–241.

[14] BIBBY, B. M., AND SØRENSEN, M. Simplified estimating functions for diffusions models with a high-dimensional parameter. *Scand. J. Statist. 28* (2001), 99–112.

[15] BIBBY, B. M., AND SØRENSEN, M. Diffusion-type models with given marginal distribution and autocorrelation function. Tech. rep., Preprint,University of Copenhagen, 2003.

[16] BLÆSILD, P., AND SØRENSEN, M. '**hyp**' a computer program for analyzing data by means of the hyperbolic distribution. *Research Report 248, Department of Theoretical Statistics, University of Arhus.* (1992).

[17] BOLLERSLEV, T. A conditionally heteroskedastic time series model for speculative prices and rates of return. *The Review of Economics and Statistics. 69* (1987), 542–547.

[18] BONDESSON, L. On simulation from infinitely divisible distributions. *Adv. in Appl. Probab. 14* (1982).

[19] BONDESSON, L. *Generalized gamma convolutions and related classes of distributions and densities.* Lecture notes in statistics. 76. Springer-Verlag, 1992.

[20] BROCKWELL, P. J., AND DAVIS, R. A. *Introduction to time series and forecasting.* Springer-Verleg, 1996.

[21] CASELLA, G., AND BERGER, R. L. Estimation with selected binomial information or do you really believe that Dave Winfield is batting .471? *J. Amer. Stat. Assoc. 89* (1994), 1080–1090.

[22] COX, D. R. Statistical analysis of time series: some recent developments. *Scand. J. Statist. 8* (1981), 93–115.

[23] COX, D. R., AND MILLER, H. D. *The theory of stochastic processes.* William Clowes and Sons. London, 1965.

[24] COX, J., INGERSOLL, J., AND ROSS, S. A theory of the term structure of interest rates. *Econometrica. 53*, 2 (1985), 385–408.

[25] CRAMÉR, H., AND LEADBETTER, M. R. *Stationary and related stochastic processes. Sample function properties and their applications.* John Wiley & Sons Inc., New York, 1967.

[26] DEMPSTER, A. P., LAIRD, N. M., AND RUBIN, D. B. Maximum likelihood from incomplete data via the EM algorithm. *J. R. Statist. Soc. B. 39*, 1 (1977), 1–38. With discussion.

[27] DOOB, J. L. *Stochastic processes.* John Wiley & Sons, New York, 1953.

[28] DOORNIK, J. A. *Object-Oriented Matrix Programming Using Ox*, 3 ed. Timberlake Consultants Press and Oxford, 2002.

[29] EBERLEIN, E. Application of generalized hyperbolic Lévy motions to finance. In *Lévy processes. Theory and Applications*, O. E. Barndorff-Nielsen, T. Mikosch, and S. I. Resnick, Eds. Birkhauser, 2001, pp. 319–337.

[30] EBERLEIN, E., AND HAMMERSTEIN, E. A. Generalized hyperbolic and inverse Gaussian distributions: Limiting cases and approximation of processes. *Preprint 80. Freiburger Zentrum für Datenanalyse und Modellbildung* (2002).

[31] ELLIOTT, R., AGGOUN, L., AND MOORE, J. *Hidden Markov Models: Estimation and Control.* Springer-Verlag, 1994.

[32] ENGLE, R. F. Autoregressive conditional heteroscedasticity with estimates of the variance of united kingdom inflation. *Econometrica. 50*, 4 (1982), 987–1008.

[33] FAMA, E. F. The behaviour of stock market prices. *Journal of Business 38* (1965), 34–105.

[34] FELLER, W. *An introduction to probability theory and its applications.*, vol. II. Wiley, 1971.

[35] FERGUSON, T. A Bayesian analysis of some nonparametric problems. *Ann. Statist. 1* (1973), 209–230.

[36] FERGUSON, T. Prior distributions on spaces of probability measures. *Ann. Statist. 2* (1974), 615–629.

[37] FERGUSON, T. S., AND KLASS, M. J. A representation of independent increment processes without Gaussian components. *Ann. Math. Statist. 43* (1972), 1634–1643.

[38] GALLANT, A. R., HSIEH, D., AND TAUCHEN, G. Estimation of stochastic volatility models with diagnostic. *Journal of Econometrics 81* (1997), 159–192.

[39] GAVER, D. P., AND LEWIS, P. A. W. First-order autoregressive gamma sequences and point processes. *Adv. in Appl. Probab. 12* (1980), 727–745.

[40] GIKHMAN, I. I., AND SKOROKHOD, A. V. *Introduction to the theory of random processes.* Dover, 1996. [Russian original 1965].

[41] GÖING-JAESCHKE, A., AND YOR, M. A survey and some generalizations of Bessel processes. *Research Report. Department of Mathematical Sciences, University of Aarhus.* (2002).

[42] HART, J. D. On the marginal distribution of a first order autoregressive process. *Statist. Probab. Lett. 2* (1984), 105–109.

[43] IBRAGIMOV, I., AND ROZANOV, I. *Gaussian Random Processes.* Springer Verlag, 1978.

[44] JACOBS, P. A., AND LEWIS, P. A. W. A mixed autoregressive-moving average exponential sequence and point process (EARMA(1, 1)). *Adv. in Appl. Probab. 9* (1977), 87–104.

[45] JACOD, J., AND SHIRYAEV, A. N. *Limit theorems for stochastic process.* Springer-Verlag, 1987.

[46] JAMES, L. F., LIJOI, A., AND PRÜNSTER, I. Predictable alternatives for Dirichlet processes. *Submitted manuscript* (2002).

[47] JENSEN, M. B., AND LUNDE, A. The NIG-S&ARCH model: A fat tailed, stochastic, and autoregressive conditional heteroskedastic volatility model. *Econometrics Journal 4* (2001), 319–342.

[48] JOE, H. Time series models with univariate margins in the convolution-closed infinitely divisible class. *J. Appl. Probab. 33* (1996), 664–677.

[49] JØRGENSEN, B. *Statistical properties of the generalized inverse Gaussian distribution.* Lecture notes in statistics, Vol. 9. Springer-Verlag, 1982.

[50] JØRGENSEN, B., AND SONG, P. Stationary time series models with exponentials dispersion model margins. *J. Appl. Probab. 35* (1998), 78–92.

[51] JUREK, Z. J., AND MASON, J. D. *Operator-limit distributions in probability theory.* Wiley, New York, 1993.

[52] KALLENBERG, O. *Foundations of Modern Probability.*, 2 ed. Springer, 2001.

[53] KARLIN, S., AND TAYLOR, H. M. *A second course in stochastic processes.* Academic Press, 1981.

[54] KENT, J. Time-reversible diffusions. *Adv. in Appl. Probab. 10* (1978), 819–835.

[55] KÜCHLER, U., AND SØRENSEN, M. Exponential families of stochastic processes and Lévy processes. *J. Statist. Plann. Inference. 39* (1994), 211–237.

[56] KÜCHLER, U., AND SØRENSEN, M. *Exponential Families of Stochastic Processes.* Springer-Verlag New York, 1997.

[57] LAVINE, M. Some aspects of Pólya tree distributions for statistical modelling. *Ann. Statist. 20* (1992), 1222–1235.

[58] LAWRANCE, A. J. The innovation distribution of a gamma distributed autoregressive process. *Scand. J. Statist. 9* (1982), 234–236.

[59] LAWRANCE, A. J. Directionality and reversivility in time series. *Internat. Statist. Rev. 59* (1991), 67–79.

[60] LAWRANCE, A. J., AND LEWIS, P. A. W. An exponential moving-average sequence and point process (EMA1). *J. Appl. Probab. 14* (1977), 98–113.

[61] LAWRANCE, A. J., AND LEWIS, P. A. W. The exponential autoregressive-moving average EARMA(p, q) process. *J. Roy. Statist. Soc. Ser. B 42* (1980), 150–161.

[62] LENK, P. J. The logistic normal distribution for Bayesian nonparametric, predictive densities. *J. Amer. Statist. Asoc. 83*, 402 (1988), 509–516.

[63] LEONARD, T. Density estimation, stochastic processes, and prior information. *J. R. Statist. Soc. B. 40* (1978), 113–146.

[64] LÉVY, P. Sur les intégrales dont les éléments sont des variables aléatoires indépendantes. *Ann. Scuola Norm. Sup. Pisa 2*, 3 (1934), 337–336.

[65] LIU, J. S., WONG, W. H., AND KONG, A. Covariance structure of the Gibbs sampler with applications to the comparisons of estimation and augmentation schemes. *Biometrika 81* (1994), 27–40.

[66] LO, A. Y. On a class of Bayesian nonparametric estimates: I, density estimates. *Ann. Statist. 12* (1984), 351–357.

[67] LOUIS, T. A. Finding the observed information matrix when using the EM algorithm. 226–233.

[68] LUKACS, E. A characterization of stable processes. *J. Appl. Probab. 6* (1969), 409–418.

[69] MANDELBROT, B. B. New methods in statistical economics. *Journal of Political Economy 71* (1963), 421–440.

[70] MAULDIN, R., SUDDERTH, W., AND WILLIAMS, S. Pólya trees and random distributions. *Ann. Statist. 20* (1992), 1203–1221.

[71] MCDONALD, I. L., AND ZUCCHINI, W. *Hidden Markov and other models for discrete-valued time series.* Chapman and Hall/CRC Press. London/Baton Rouge, 1997.

[72] McKenzie, E. Autoregressive moving-average processes with negative-binomial and geometric marginal distributions. *Adv. in Appl. Probab. 18* (1986), 679–705.

[73] McKenzie, E. Some ARMA models for dependent sequences of Poisson counts. *Adv. in Appl. Probab. 20* (1988), 822–835.

[74] Mena, R. H., and Walker, S. G. Distribution of the innovation part of a self-decomposable random variable. *Accepted for publication. Journal of Statistical Computation and Simulation* (2003).

[75] Mittnik, S., and Rachev, S. Modeling asset returns with alternative stable distributions. *Econometric Reviews 12*, 3 (1993), 261–330.

[76] Muliere, P., and Walker, S. G. A Bayesian non-parametric approach to survival analysis using Pólya trees. *Scand. J. Statist. 24* (1997), 331–340.

[77] Nelson, D. B. Arch models as diffusion approximations. *Journal of Econometrics 45* (1990), 7–38.

[78] Pitt, M. K., Chatfield, C., and Walker, S. G. Constructing first order autoregressive models via latent processes. *Scand. J. Statist. 29* (2002), 657–663.

[79] Pitt, M. K., and Walker, S. G. A stationary ARCH(1) time series model. *Submitted manuscript* (2003).

[80] Press, W., Teukolsky, S., Vetterling, W., and Flannery, B. *Numerical Recipes in C.* Cambridge University Press, 1992.

[81] Revuz, D., and Yor, M. *Continuous Martingales and Brownian motion*, 3 ed. Springer, 1998.

[82] Robert, C. P., and Casella, G. *Monte Carlo statistical methods*, third ed. Springer-Verlag, 2002.

[83] Rogers, L. C. G. A guided tour through excursions. *Bulletin of the London Mathematical Society.*, 21 (1989), 305–341.

[84] Rogers, L. C. G., and Williams, D. *Diffusions, Markov processes and martingales*, second ed., vol. I. Foundations. Cambridge University Press, 2000.

[85] ROGERS, L. C. G., AND WILLIAMS, D. *Diffusions, Markov processes and martingales.*, second ed., vol. II. Itô Calculus. Cambridge University Press, 2000.

[86] ROSIŃSKI, J. Series representations of Lévy processes from the perspective of point processes. In *Lévy processes.* Birkhäuser Boston, Boston, MA, 2001, pp. 401–415.

[87] RUBINSTEIN, R. Y. *Simulation and the Monte Carlo method.* John Wiley & Sons Inc., New York, 1981. Wiley Series in Probability and Mathematical Statistics.

[88] RYDBERG, T. H. Generalized hyperbolic diffusion processes with applications in finance. *Mathemathical Finance 3* (1999), 183–201.

[89] SARNO, E. Dependence structures of Pólya tree autoregressive models. *Statistica LVIII*, 3 (1998), 363–373.

[90] SATO, K. I. Self-similar processes with independent increments. *Probab. Theory Related Fields. 89* (1991), 285–300.

[91] SATO, K. I. *Lévy processes and infinitely divisible distributions.* Cambridge University Press, 1999.

[92] SATO, K. I. Subordination and self-decomposability. *Statist. Probab. Lett. 54* (2001), 317–324.

[93] SCHERVISH, M. J. *Theory of statistics.* Springer-Verlag, 1996.

[94] SHIGA, T., AND WATANABE, S. Bessel diffusions as a one-parameter family of diffusion processes. *Z. Wahrsch. Verw. Gebiete 27* (1973), 37–46.

[95] SILVERSTEIN, M. L. *Symmetric Markov processes.* Springer-Verlag, Berlin, 1974. Lecture Notes in Mathematics, Vol. 426.

[96] TANNER, M. A., AND WONG, W. H. The calculation of posterior distributions by data augmentation. *J. Amer. Statist. Assoc. 82* (1987), 528–550. With discussion and with a reply by the authors.

[97] THORBURN, D. A Bayesian approach to density estimation. *Biometrika 73* (1986), 65–76.

[98] TUDOR, C. *Procesos estocásticos.*, 2 ed. Sociedad matemática mexicana. Aportaciones matemáticas, 1997. In Spanish.

[99] VERVAAT, W. On a stochastic difference equation and a representation of nonnegative infinitely divisible random variables. *Adv. in Appl. Probab. 11* (1979), 750–783.

[100] WALKER, S. G. An EM algoritm for nonlinear random effects models. *Biometrics 52* (1996), 934–944.

[101] WALKER, S. G. Comment on the paper [13].

[102] WALKER, S. G., DAMIEN, P., LAUD, P. W., AND SMITH, A. F. M. Bayesian nonparametric inference for random distributions and related functions (with discussion). *J. R. Statist. Soc. B. 61* (1999), 485–527.

[103] WEI, G. C., AND TANNER, M. A. A Monte Carlo implementation of the EM algorithm and the poor man's data augmentation algorithms. *J. Amer. Stat. Assoc. 85* (1990), 699–704.

[104] WEISS, G. Time-reversibility of linear stochastic processes. *J. Appl. Probab. 12* (1975), 831–836.

[105] WOLFE, S. J. On the unimodality of L functions. *Ann. Math. Statist. 42* (1971), 912–918.

[106] WOLFE, S. J. On the unimodality of multivariate symmetric distribution functions of class L. *J. Multivar. Anal. 8* (1978), 141–145.

[107] WOLFE, S. J. On a continuous analogue of the stochastic difference equation $X_n = \rho X_{n-1} + B_n$. *Stoch. Proc. Appl. 12* (1982), 301–312.

[108] YACINE, A. S. Nonparametric pricing of interest rate derivative securities. *Econometrica 64*, 3 (1996), 527–560.

[109] YACINE, A. S., AND MYKLAND, P. A. The effect of random and discrete sampling when estimating continuous-time diffusions. *Econometrica 71*, 2 (2003), 483–549.

[110] YAMAZATO, M. Unimodality of infinitely divisible distribution functions of class L. *Ann. Probab. 6* (1978), 523–531.